

Epistemología bayesiana

Michael G. Titelbaum

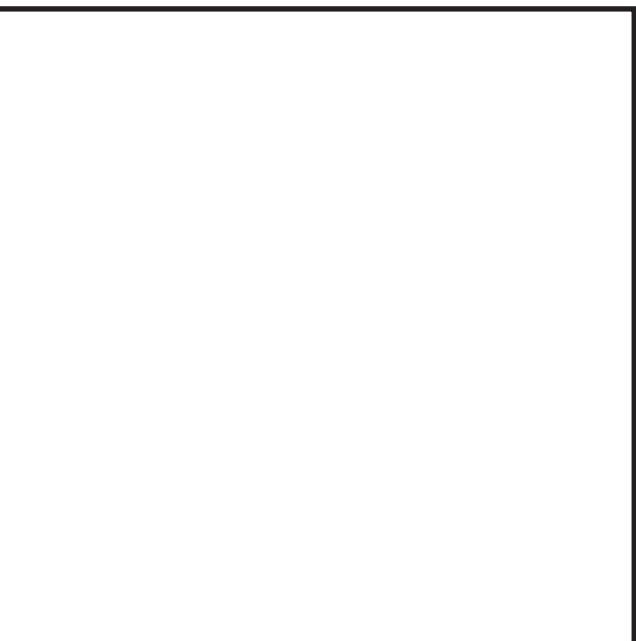


UNIVERSIDAD
AUTÓNOMA DE
AGUASCALIENTES

Epistemología bayesiana

Michael G. Titelbaum

Marc Jiménez Rolland
Traducción



Epistemología bayesiana

Michael G. Titelbaum

Marc Jiménez Rolland
Traducción



UNIVERSIDAD AUTÓNOMA
DE AGUASCALIENTES

Epistemología bayesiana

Primera edición 2024 (versión electrónica)

© Universidad Autónoma de Aguascalientes
Av. Universidad 940
Ciudad Universitaria, 20100
Aguascalientes, México

© Michael G. Titelbaum

© Marc Jiménez Rolland
Traducción

Este texto es una traducción del capítulo “Precise Credences”, de Michael G. Titelbaum, que se incluye en el *Open Handbook of Formal Epistemology* (páginas 1 a 55), editado por Richard Pettigrew y Jonathan Weisberg en 2019, bajo el sello de la PhilPapers Foundation. Fue publicado con una Licencia Internacional Creative Commons Attribution-NonCommercial-NoDerivatives 4.0. La presente traducción se publica con autorización del autor y de los editores.

ISBN 978-607-8972-86-9

Hecho en México / *Made in Mexico*



editorial.uaa.mx



libros.uaa.mx



libreriavirtual.uaa.mx



revistas.uaa.mx

Índice

La introducción de credibilidades	9
Restricciones racionales sobre la credibilidad	15
Los axiomas de la probabilidad	15
La Fórmula de Proporción	19
Actualizar por Condicionización	24
Condicionización de Jeffrey	28
Requisitos racionales adicionales	31
Principios de inferencia directa	33
Aditividad Contable	36
 Aplicaciones de la credibilidad	 41
La teoría de la confirmación	41
La teoría de la decisión	46
Otras aplicaciones	49
 Argumentos a favor de restricciones credales	 51
Argumentos de Teorema de Representación	51
Argumentos de Libro Holandés	53
Argumentos de Precisión	57

Argumentos en contra de restricciones credales	63
El Problema de la Omnisciencia Lógica	63
El Problema de la Vieja Evidencia	67
Pérdida de Memoria y Sensibilidad al Contexto	71
Otros formalismos de la convicción	75
Apéndice	77
Referencias	81

La introducción de credibilidades

Estoy más convencido de que iré a la oficina mañana de lo que lo estoy de que no iré. No estoy *seguro* de que iré, y ni siquiera he llegado al punto de *creer* que iré: es verano, no tengo clases que impartir o estudiantes con quienes reunirme, podría despertar mañana y decidir que no vale la pena el esfuerzo. Pero tengo mayor convicción en que iré de la que tengo en que no iré. Si tuviera que poner a mi convicción en una escala de 0 a 100, la pondría en algún lugar por encima de 50.

Las credibilidades (*credences*)* son grados numéricos de convicción. Aunque podrían expresarse como porcentajes –entre 0 y 100, incluidos éstos– se ha vuelto habitual medirlas en una escala del 0 al 1. A las credibilidades también se les llama a menudo “grados de creencia”, aunque ese nombre podría tener la connotación de que son una especie de creencias cualitativas ordinarias.

Es mejor pensar en la credibilidad no como una clase de creencia cualitativa, sino más bien como un miembro de la misma familia que la creencia cualitativa. Esa familia –la familia de las actitudes doxásticas– también incluye la certeza, la incredulidad, la

* Nota del traductor: Al igual que el término inglés ‘*credence*’, la palabra ‘credibilidad’ tiene un uso común en castellano. Sin embargo, en ambos idiomas su uso como sustantivo *contable* en el presente contexto –al hablar de credibilidades de agentes hacia afirmaciones específicas– es relativamente reciente. En su libro *Fundamentals of Bayesian Epistemology* (Oxford University Press), Michael Titelbaum (2022, p. 22, n. 17) encuentra un precedente del uso que predomina en este texto en el artículo de Rudolf Carnap “The Aim of Inductive Logic”: “El concepto de probabilidad en el sentido de grado de creencia *real* es un concepto psicológico [...]. Usaré para este concepto psicológico el término técnico ‘grado de credibilidad’ o sólo ‘credibilidad’ para abreviar” (1962: 305).

suspensión de la creencia y probablemente también la convicción comparativa. Los miembros de esta familia tienen una amplia variedad de rasgos comunes. Por ejemplo, tendemos a pensar en las credibilidades como teniendo los mismos tipos de *objetos* que las creencias plenas (*outright belief*). Muchos autores consideran que estos objetos son proposiciones, por ello clasifican a las credibilidades y a las creencias como actitudes proposicionales. Seguiré esa tendencia aquí, pero si crees que las creencias se adoptan hacia alguna otra cosa que proposiciones (¿oraciones, quizá?), entonces te inclinarás a tener el mismo punto de vista sobre las credibilidades.

Se desarrolló la teoría de las credibilidades para tratar varios problemas filosóficos. Uno era la interpretación adecuada de locuciones sobre “probabilidad”. Si digo “la probabilidad de que vaya a la oficina mañana es mayor que 50%”, ¿qué significa eso?, ¿y cuáles son las condiciones de verdad de mi preferencia? Se han ofrecido y defendido diversas interpretaciones de la probabilidad (discutiremos algunas en el apartado “Principios de inferencia directa”, en las páginas 33 a 36), y no es claro que todo uso del término “probabilidad” deba interpretarse de la misma manera. Pero una sugerencia prominente, la “interpretación subjetiva de la probabilidad”, es que los enunciados sobre probabilidad expresan el grado de convicción del hablante en una proposición. De modo que mi preferencia expresa una convicción superior a 0.5 en que iré a la oficina.

Incluso si los enunciados sobre “probabilidad” rara vez —o nunca— expresan los grados de convicción de una agente, tales grados de convicción pueden aun así existir y tener trabajo filosófico por hacer. Los grados de creencia juegan un papel importante en la teoría de la decisión tradicional, la aproximación formal clásica a la elección racional (más al respecto en las páginas 46 a 48). Las credibilidades también aparecen en la teoría bayesiana de la confirmación (páginas 41 a 46), una explicación del apoyo evidencial que rivaliza con otras aproximaciones estadísticas como el frecuentismo y el verosimilitudismo (*likelihoodism*). Y pueden aplicarse a temas adicionales como el coherentismo, la inferencia a la mejor explicación y la epistemología social (páginas 49 y 50).

Así que si aceptamos que las credibilidades existen, ¿en qué consiste exactamente poseer una? En línea con las aproximaciones conductistas en la psicología de su época, De Finetti (1937/1964) *definió* el grado de creencia asignado a un suceso por un individuo como la proporción en la que apostaría a que ocurriría (más al

respecto en las páginas 46 a 48). Pero, como es típico con el operacionalismo, esta definición encontró problemas cuando, por ejemplo, una agente mostraba conducta de apuestas inconstante a través del tiempo y por ende era difícil asignarle una credibilidad particular. Hoy en día podemos conceder que una agente con un grado particular de creencia, *si es racional*, mostrará una conducta de apuestas particular (Christensen, 2004). Pero también tendemos a pensar en esta conexión normativa menos como una *definición* de credibilidad y más como un aspecto de en qué consiste poseer un grado de convicción. Tal como nuestra explicación de la creencia cualitativa ha progresado más allá del conductismo hacia un funcionalismo más amplio, pensamos en la credibilidad como un estado mental multifacético con conexiones descriptivas y normativas hacia una amplia variedad de conductas y otras actitudes.

Además de sus conexiones con deseos, intenciones y decisiones contempladas en la teoría de la acción y en la teoría de la decisión, las credibilidades se conectan con otras variedades de actitudes doxásticas (por no mencionar emociones, sensaciones y recuerdos). Si la convicción comparativa es un tipo distinto de estado mental, claramente se conecta con la credibilidad: estoy más convencido de P que de Q sólo en caso de que mi credibilidad en P sea más alta que mi credibilidad en Q . En cuanto a las actitudes cualitativas, a menudo se identifica a la certeza con la credibilidad 1 en una proposición (aunque véase la discusión en las páginas 36 a 39). Debe haber también vínculos entre credibilidad y creencia plena: si creo que P , mi credibilidad en P debe ser mayor que mi credibilidad en $\neg P$.

¿Podemos encontrar una conexión completamente general entre credibilidad y creencia plena? Algunos autores (*e.g.*, Holton, 2014) sostienen que, en la medida en que hay credibilidades, poseer una credibilidad x en P es simplemente tener una creencia plena en que la probabilidad de P es x . Aun así, es difícil encontrar un único concepto de probabilidad que se aplique a todas las proposiciones a las que una agente podría asignarles un grado de creencia. Y parece que agentes (como los niños) pueden estar más o menos convencidos de proposiciones sin poseer un concepto de probabilidad. Además, sin importar qué concepto de probabilidadelijamos, parece concebible que una agente adopte un grado de convicción en la proposición de que P tiene x probabilidad. (Veremos una dificultad técnica adicional con la teoría de la credibilidad-como-creencia-plena en las páginas 23 a 24). La mayoría de

los teóricos en nuestros días sostienen que el valor numérico de una credibilidad es un atributo de la actitud adoptada *hacia* una proposición, no parte del *contenido* de la proposición hacia la que se adopta la actitud.¹

En la otra dirección, la Tesis lockeana² considera a la creencia plena simplemente como una credibilidad por encima de un umbral particular. El umbral de credibilidad usualmente es menor que 1 (la creencia no requiere ser certeza), pero bastante por encima de $\frac{1}{2}$, y puede depender de parámetros contextuales. La principal objeción a la Tesis lockeana es que uno puede describir distribuciones racionalmente aceptables de credibilidad que, por medio de la Tesis, generen patrones racionalmente inaceptables de creencia. En la Paradoja de la lotería (Kyburg, 1961), una agente asigna a cada boleto de una lotería una baja credibilidad de que ganará, mientras que asigna una alta credibilidad (quizá certeza) a que algún boleto ganará. Para cualquier umbral lockeano menor que 1, podemos acomodar los números de modo que la agente acabe creyendo de cada boleto que perderá, mientras cree que algún boleto ganará —un conjunto total de creencias lógicamente inconsistente—. De manera similar, en la Paradoja del prefacio (Makinson, 1965), una autora tiene elevada convicción en cada afirmación que se hace en su libro mientras también está convencida de que al menos una de esas afirmaciones es falsa. Por medio de la Tesis lockeana esto se convierte en la creencia en cada miembro de una conjunción unida con la incredulidad sobre esa conjunción.

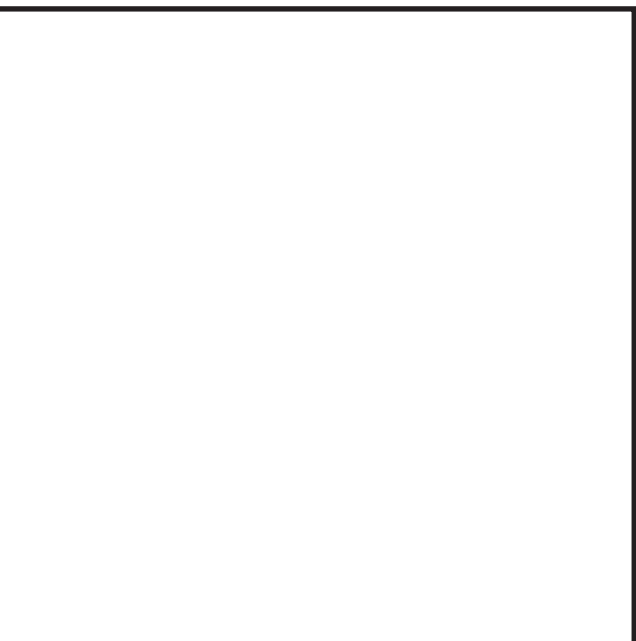
¿Cómo, entonces, relacionar credibilidad con creencia plena en general? La posibilidad más radical es negar ya sea la existencia de creencias o bien la existencia de credibilidades. De manera más conservadora, uno podría ofrecer una reducción de una categoría a la otra, o al menos un principio descriptivo de sobreviniencia. Alternativamente, uno podría conceder que mientras las creencias y las credibilidades aparecen en gran variedad de configuraciones en las agentes *reales*, los principios normativos especifican cómo se alinean en una agente *racional*. El consenso en la actualidad es que se requeriría algo más allá de la simple Tesis lockeana para hacer que cualquiera de estas aproximaciones funcione; pueden encontrarse intentos recientes de articular princi-

1 Moss (2018) considera que el valor numérico es parte del contenido de una credibilidad, pero asume que los objetos credales son más complicados que simples proposiciones.

2 Locke (1689, L. IV). Véase también Foley (1993) para discusión.

pios de creencia-credibilidad en Leitgeb (2017), Douven (2012) y Lin y Kelly (2012).

Por otra parte, se podría conceder que las creencias y las credibilidades son ambas clases genuinas de estados mentales que una agente puede poseer, que hay algunas maneras en las que interactúan (o en las que interactúan si se es racional), pero que no están disponibles principios generales sistemáticos. Aunque los realistas robustos sobre creencias y credibilidades pueden adoptar esta actitud, es especialmente atractiva para los teóricos que leen las adscripciones de creencia y credibilidad como modelos convenientes y simplificadores de un sistema cognitivo altamente complejo. El modelo de creencia y el modelo de credibilidad son, cada uno por su cuenta, efectivos y eficientes en diferentes circunstancias y pueden aplicarse con diferentes fines. En ese caso, no sería sorprendente si una traducción universal de uno al otro no estuviese disponible.



Restricciones racionales sobre la credibilidad

Una vez que entendemos qué es una credibilidad, la siguiente pregunta es en qué consiste que un conjunto de credibilidades sea racional.

Los axiomas de la probabilidad

Las normas de la credibilidad racional más generalmente aceptadas son los axiomas de Kolmogorov (1933/1950). Supóngase que tenemos un lenguaje $\mathcal{L}$ de proposiciones, el cual comienza con un conjunto finito de proposiciones atómicas y luego las cierra bajo las conectivas veritativo-funcionales estándar. Definase una función c sobre $\mathcal{L}$ con valores en los números reales para representar los valores de credibilidad que una agente asigna a las proposiciones en $\mathcal{L}$.³ Los valores precisos en números reales que c asigna a cada proposición son las “credibilidades precisas” en el título de esta entrada;* discutiré aproximaciones formales alternativas en las páginas 75 y 76.

En este arreglo, los axiomas de Kolmogorov se vuelven los siguientes.

NO-NEGATIVIDAD. Para cualquier $X \in \mathcal{L}$, $c(X) \geq 0$.

NORMALIDAD. Para cualquier tautología $\top \in \mathcal{L}$, $c(\top) = 1$.

3 Aunque consideraré lenguajes que contienen proposiciones, otros autores describen a las credibilidades como distribuidas sobre oraciones, o conjuntos de mundos posibles, o conjuntos de sucesos, etc.

* Nota del traductor: Se refiere al título original del texto “Precise Credences”.

ADITIVIDAD FINITA. Para cualesquiera $X, Y \in \mathcal{L}$ mutuamente excluyentes, $c(X \vee Y) = c(X) + c(Y)$.

Los matemáticos a menudo llaman a éstos *los axiomas de la probabilidad* y denominan a cualquier distribución que los satisfaga una *función de probabilidad*. El probabilismo es la posición que sostiene que las credibilidades racionales forman una función de probabilidad; en otras palabras, las credibilidades racionales satisfacen los axiomas de Kolmogorov.⁴

Los axiomas de la probabilidad establecen $0 \leq c(X) \leq 1$ para todo $X \in \mathcal{L}$. El probabilismo tiene también como consecuencias lógicas varias restricciones intuitivas sobre las credibilidades racionales. He aquí un ejemplo.

- Para cualquier $X \in \mathcal{L}$, $c(\neg X) = 1 - c(X)$.

Supón que le asignas una alta convicción a que ha ocurrido el calentamiento global antropogénico. Esta restricción exige que le asignes una baja convicción a que no ha ocurrido el calentamiento antropogénico. Y en caso de que te convencieras aun más de que ha ocurrido el calentamiento antropogénico, esta restricción requerirá que tu convicción en la negación de esa proposición disminuya.

Algunas otras restricciones intuitivas que se siguen de los axiomas de Kolmogorov son las siguientes.

- Para cualquier contradicción $\perp \in \mathcal{L}$, $c(\perp) = 0$.
- Para cualesquiera $X, Y \in \mathcal{L}$ (sean mutuamente excluyentes o no), $c(X \vee Y) = c(X) + c(Y) - c(X \& Y)$.
- Para cualesquiera $X, Y \in \mathcal{L}$ lógicamente equivalentes, $c(X) = c(Y)$.

⁴ A menudo se describe el probabilismo como la doctrina que sostiene que las agentes racionales tienen credibilidades que satisfacen los axiomas de la probabilidad o (si eso se considera demasiado poco realista) que las agentes idealmente racionales tienen credibilidades probabilísticas. En ambas formulaciones se hace a las agentes (reales o ideales) el objetivo de evaluación. Estrictamente hablando, yo prefiero evaluar la racionalidad de las credibilidades (o conjuntos de credibilidades), en lugar de la de las agentes. Pero por facilidad de expresión, en su mayoría trataré las dos como intercambiables aquí.

- Para cualquier conjunto finito de $X_1, \dots, X_n \in \mathcal{L}$ mutuamente excluyentes, $c(X_1 \vee \dots \vee X_n) = c(X_1) + \dots + c(X_n)$.

La restricción de la última viñeta tiene una importante consecuencia cuando una agente considera una *partición* —un conjunto de proposiciones cuyos miembros son mutuamente excluyentes y conjuntamente exhaustivos—. Debido a que la disyunción de los elementos de una partición es una tautología, el probabilismo exige que la suma de las credibilidades que se le asignan a los elementos de una partición sea igual a 1.

Una consecuencia adicional importante del probabilismo es que las credibilidades son robustamente extensionales. Si una agente está segura de que dos proposiciones X y Y tienen el mismo valor de verdad (es decir, si $c(X \equiv Y) = 1$), entonces con el fin de calcular credibilidades, X y Y podrían también ser lógicamente equivalentes. Por ejemplo, cualquier desigualdad o ecuación de credibilidad en la que aparece X seguiría siendo verdadera si cualquiera de sus X s fuese reemplazada con Y s. Cualquier diferencia en significado, perfil modal, etc., es irrelevante para la probabilidad una vez que se establece que sus valores de verdad son idénticos.

Podemos ilustrar el probabilismo con el ejemplo de Kyburg de la lotería de la página 12. Dada una lotería con, digamos, 100 boletos, introduce un lenguaje cuyas proposiciones atómicas son G_i hasta G_{100} (con G_i indicando que el boleto i gana la lotería). Si la lotería no está amañada, una agente podría asignar $c(G_i) = 1/100$ para cada G_i . A partir de nuestra primera consecuencia intuitiva de los axiomas de la probabilidad, tenemos entonces que $c(\neg G_i) = 99/100$; la agente está altamente convencida sobre cada boleto de que no ganará. Sin embargo, asumiendo que no más de un boleto puede ganar, nuestra consecuencia intuitiva final enlistada arriba produce:

$$c(G_1 \vee \dots \vee G_{100}) = c(G_1) + \dots + c(G_{100}) = 1. \quad (1)$$

Así que nuestra agente está segura de que algún boleto ganará, tal como tendría que ser intuitivamente.⁵

5 Nota que nada de esto resuelve la Paradoja de la lotería, la cual introduce creencias completas en la imagen de la lotería. Mi objetivo es sólo ilustrar cómo el probabilismo es compatible con, y apoya a, una explicación natural de las credibilidades racionales en el caso de la lotería. Una ilustración similar podría darse del ejemplo del Prefacio de Makinson.

Aunque las demostraciones en el cálculo de probabilidades habitualmente proceden a partir de los axiomas de Kolmogorov, a menudo es más sencillo resolver problemas prácticos trabajando con descripciones de estado. Defínase una literal como una proposición atómica de $\mathcal{L}$ o su negación, luego defínase una descripción de estado en $\mathcal{L}$ como una conjunción máximamente consistente de sus literales. Cualquier $X \in \mathcal{L}$ no contradictoria tiene una forma normal disyuntiva única, una disyunción de descripciones de estado lógicamente equivalente a X .⁶

Carnap (1950) hace uso reiterado del hecho de que una distribución c sobre $\mathcal{L}$ satisface los axiomas de la probabilidad sólo en caso de que asigne: (1) valores no negativos a las descripciones de estado de $\mathcal{L}$ que sumen 1; (2) para cualquier X no contradictoria, un valor igual a la suma de los valores asignados a la descripción de estado en la forma normal disyuntiva de X ; y (3) un valor de 0 a cualquier proposición contradictoria.⁷

Este resultado es útil de dos maneras. Primero, podemos caracterizar completamente cualquier distribución de probabilidad sobre $\mathcal{L}$ al especificar los valores que asigna a las descripciones de estado de $\mathcal{L}$. Segundo, dada información parcial sobre una distribución de probabilidad, podemos determinar lo que esta información dice sobre los valores asignados a descripciones de estado, y luego determinar desde ahí los valores de (o las restricciones sobre los valores de) otras proposiciones.

Por ejemplo, supón que te digo que Bob está seguro de que $P \supset Q$ y tiene el doble de convicción en P con respecto a $\neg P$. Se sigue inmediatamente que la convicción de Bob en $\neg Q$ es menor o igual a $1/3$. ¿Por qué? Bien, la forma normal disyuntiva equivalente de $\neg Q$ es $(P \& \neg Q) \vee (\neg P \& \neg Q)$. Puesto que Bob está seguro de que $P \supset Q$, el primer disyunto recibe credibilidad de 0, de modo que para Bob $c(\neg Q) = c(\neg P \& \neg Q)$. Pero puesto que $c(P) + c(\neg P) = 1$ y $c(P) = 2 \cdot c(\neg P)$, tenemos que $c(\neg P) = 1/3$. La forma normal disyuntiva equivalente de $\neg P$ es $(\neg P \& Q) \vee (\neg P \& \neg Q)$. Por No-Negativi-

6 Para hacer la forma normal disyuntiva única, requerimos que las literales aparezcan en una descripción de estado en algún orden canónico (quizá alfabético, si se designa a las proposiciones por medio de letras) y luego requerimos que las descripciones de estado aparezcan en formas normales disyuntivas en un orden canónico también.

7 Nunca he sido capaz de descubrir si éste fue un resultado original de Carnap o no. ¡Agradecería sinceramente cualquier correo electrónico sobre su origen histórico!

dad, la credibilidad de Bob en el primer disyunto debe ser mayor o igual a 0, de modo que el segundo disyunto recibe una credibilidad menor o igual a $1/3$.⁸

Finalmente, con la noción de una función de probabilidad a la mano podemos definir la noción de una *expectativa* (*expectation*). Supóngase que tenemos una cantidad numérica para la cual muchos valores son posibles. Para calcular la expectativa de una agente para esa cantidad, multiplicamos cada valor por la credibilidad de la agente en que esa cantidad asumirá ese valor, luego sumamos todos los valores disponibles. Por ejemplo, si tengo una convicción de 10% en que iré a mi oficina dos días esta semana, estoy 60% convencido de que iré sólo un día y 30% convencido de que no iré en absoluto, entonces mi expectativa para el número de días que iré a mi oficina esta semana es:

$$0.10 \cdot 2 \text{ días} + 0.60 \cdot 1 \text{ día} + 0.30 \cdot 0 \text{ días} = 0.8 \text{ días.} \quad (2)$$

La Fórmula de Proporción

Hasta ahora hemos discutido la credibilidad incondicional –el grado de convicción de una agente en que una proposición particular es verdadera a la luz de su entendimiento de cómo es el mundo en ese momento–. También podemos investigar la credibilidad *condicional* de una agente en la proposición X dado que $\mathcal{I}$; ésta es la credibilidad de la agente en X habiendo hecho la suposición adicional de que $\mathcal{I}$. Nota que $\mathcal{I}$ puede ser una proposición sobre la cual la agente tiene en ese momento una baja credibilidad incondicional. Al preguntar por su credibilidad en X dado que $\mathcal{I}$, le pedimos que deje de lado su opinión sobre $\mathcal{I}$ en ese momento, temporalmente añada $\mathcal{I}$ al repertorio de proposiciones que considera que son verdaderas, luego evalúe X a la luz de este conjunto aumentado de suposiciones.⁹

8 Para más sobre la teoría matemática que subyace a esta aproximación y para un procedimiento en Mathematica que resuelve muchos problemas de probabilidad una vez que se reducen a álgebra usando descripciones de estado, véase Fitelson (2008).

9 Nota que estamos discutiendo credibilidades condicionales indicativas, no subjuntivas. La suposición $\mathcal{I}$ debe *añadirse* al conjunto de suposiciones sobre el mundo de la agente en ese momento, asumiendo que el conjunto de

La credibilidad condicional de una agente en X dado que $\mathcal{I}$ se expresa como $c(X|\mathcal{I})$, y habitualmente se asume que está gobernada por la Fórmula de Proporción.

FÓRMULA DE PROPORCIÓN

Para cualesquiera $X, \mathcal{I} \in \mathcal{L}$ con $c(\mathcal{I}) > 0$, $c(X|\mathcal{I}) = \frac{c(X\&\mathcal{I})}{c(\mathcal{I})}$.

La Fórmula de Proporción puede leerse ya sea como una verdad descriptiva o como un requisito normativo. En la primera aproximación, la credibilidad condicional de una agente en X dado que $\mathcal{I}$ toma un valor particular sólo en caso de que sus credibilidades incondicionales en $X\&\mathcal{I}$ y en $\mathcal{I}$ se encuentren en esa proporción. Esta lectura es más natural si se quiere reducir un tipo de credibilidad al otro: se podría sostener que tener una credibilidad condicional es simplemente tener credibilidades incondicionales que se encuentran en una proporción específica; o se podría sostener que las credibilidades condicionales son básicas y las credibilidades incondicionales son un subconjunto propio de aquellas.¹⁰ Alternativamente, se podría ver a la credibilidad condicional como sólo otro tipo de actitud doxástica que se encuentra a la par con las credibilidades incondicionales y luego leer la Fórmula de Proporción como un requerimiento racional sobre cómo deberían alinearse credibilidades condicionales e incondicionales.¹¹

Nota que la Fórmula de Proporción, tal como la he definido, no dice nada sobre el caso en el que la agente asigna a la condición (la proposición $\mathcal{I}$) una credibilidad de 0. Volveremos a las credibilidades condicionales sobre proposiciones con credibilidad 0 en las páginas 37 a 39.

Al combinar la Fórmula de Proporción y los axiomas de Kolmogorov se obtiene la útil Ley de la Probabilidad Total.

suposiciones resultante sea consistente. La mayoría de las discusiones sobre credibilidades condicionales se ocupan de la forma indicativa. Para un tratamiento de credibilidades condicionales subjuntivas véase Joyce (1999).

10 A partir de los axiomas de Kolmogorov y de la Fórmula de Proporción, se sigue que para cualquier $X \in \mathcal{L}$, $c(X) = c(X|\top)$. Así que puede pensarse en las credibilidades incondicionales como credibilidades condicionales en las que la condición es una tautología. Véase Easwaran (2019) para más al respecto.

11 Para una discusión sobre cómo las credibilidades condicionales interactúan con las credibilidades de una agente en condicionales, véase Briggs (2019).

LEY DE LA PROBABILIDAD TOTAL

Para cualesquiera $X, \mathcal{I}_1, \dots, \mathcal{I}_n \in \mathcal{L}$ tales que $\mathcal{I}_1, \dots, \mathcal{I}_n$ forman una partición finita, $c(X) = c(X|\mathcal{I}_1) \cdot c(\mathcal{I}_1) + \dots + c(X|\mathcal{I}_n) \cdot c(\mathcal{I}_n)$.

La Ley de la Probabilidad Total calcula la credibilidad incondicional de X como una media ponderada de las credibilidades condicionales de X sobre los miembros de la partición $\mathcal{I}$, ponderados por las credibilidades incondicionales en las $\mathcal{I}$ s.¹²

Para ilustrar una vez más con nuestro escenario de la lotería, supóngase que B es la proposición de que nuestra agente se beneficiará del resultado de la lotería. Ella tiene los boletos 1 a 3, así que es seguro que se beneficiará si éstos ganan. También, su hermana tiene el último boleto (el boleto 100) y la agente tiene una convicción de $1/2$ en que su hermana compartirá las ganancias si ese boleto resulta ganador. Al aplicar la Ley de la Probabilidad Total (y recordando que G_i es la proposición de que el boleto i ganará), la credibilidad de la agente de que ella se beneficiará es

$$\begin{aligned} c(B) &= c(B|G_1) \cdot c(G_1) + c(B|G_2) \cdot c(G_2) + \\ &\quad c(B|G_3) \cdot c(G_3) + c(B|G_4) \cdot c(G_4) + \dots + \\ &\quad c(B|G_{100}) \cdot c(G_{100}) \\ &= 1 \cdot 1/100 + 1 \cdot 1/100 + 1 \cdot 1/100 + \\ &\quad 0 \cdot 1/100 + \dots + 1/2 \cdot 1/100 \\ &= 0.035. \end{aligned} \tag{3}$$

La credibilidad condicional también desempeña un papel crucial en la noción de relevancia credal. Cuando $0 < c(\mathcal{I}) < 1$, todas las siguientes desigualdades son equivalentes:

$$c(X|\mathcal{I}) > c(X), \tag{4}$$

$$c(X) > c(X|\neg\mathcal{I}), \tag{5}$$

$$c(\mathcal{I}|X) > c(\mathcal{I}), \tag{6}$$

12 Puesto de otra manera, la Ley de la Probabilidad Total exige que las credibilidades incondicionales de una agente en X sean iguales a su expectativa de su credibilidad en X condicional sobre el elemento verdadero de la partición $\mathcal{I}$.

$$c(I) > c(I | \neg X), \quad (7)$$

$$c(X \& I) > c(X) \cdot c(I). \quad (8)$$

Cuando ocurren estas desigualdades, decimos que I es positivamente relevante a X en la función de credibilidad de la agente. (Puesto que la relevancia positiva es una relación simétrica, también podemos decir que X es positivamente relevante a I). Otra manera de poner esto es que la agente considera que X y I están positivamente correlacionadas. Al reemplazar los ‘mayor-que’ con ‘menor-que’, describimos cuándo I es negativamente relevante a X (o está negativamente correlacionado con X) en las credibilidades de una agente. Por otra parte, cuando $c(X \& I) = c(X) \cdot c(I)$ (o cualquiera de las otras desigualdades de arriba se convierte en una igualdad), decimos que X es irrelevante a I para la agente, o probabilísticamente independiente de I .

Estas relaciones de relevancia son relativas a las credibilidades de una agente; reflejan cuáles proposiciones ella considera relevantes unas a otras dado su entendimiento de cómo es el mundo en ese momento. Pero también podemos temporalmente ampliar el conjunto de sus suposiciones sobre el mundo en ese momento y ver si hay algún cambio en las relaciones de relevancia. Esto nos lleva de una noción de relevancia incondicional a la relevancia condicional. I es relevante a X condicional sobre Z sólo en caso de que

$$c(X | I \& Z) > c(X | Z). \quad (9)$$

Para cada una de las desigualdades de arriba, puede darse una caracterización correspondiente de relevancia condicional al añadir Z como una condición a la expresión en cada lado.

La noción de relevancia condicional subyace a una noción crucial en la filosofía de la ciencia: el filtrado (*screening off*) de correlaciones espurias. Decimos que Z filtra una correlación espuria de X con I cuando X y I son incondicionalmente dependientes, pero las siguientes dos igualdades se sostienen:

$$c(X | I \& Z) = c(X | Z), \quad (10)$$

$$c(X | I \& \neg Z) = c(X | \neg Z). \quad (11)$$

Con otras palabras, X y Y son condicionalmente independientes tanto de Z como de $\neg Z$. En una situación de filtrado de correlaciones espurias, suponer ya sea Z o $\neg Z$ hace a la correlación entre X y Y desaparecer.¹³

Para ilustrar una aplicación de este concepto, Reichenbach (1956) argumenta que una causa común filtra una correlación espuria entre sus efectos. Supóngase que X es la proposición de que mi periódico reporta que los Yankees ganaron anoche, Y es la proposición de que tu periódico reporta que los Yankees ganaron anoche y Z es la proposición de que los Yankees de hecho ganaron. Por otra parte, mientras ignoro Z , sería racional que yo tratara a X como relevante a Y . X proporciona información sobre Z , por ende también proporciona información sobre Y . Pero una vez que se determina el valor de verdad de Z , X y Y pierden la capacidad de decir algo una acerca de la otra; X y Y se vuelven independientes condicionales sobre cualquier suposición acerca de Z . Así, Z filtrará la correlación espuria entre X y Y en mi función de credibilidad.

Una causa próxima también filtra la correlación espuria de su efecto con una causa distante. (Imagina que Y enuncia el marcador final del juego de los Yankees anoche, Z es la proposición de que los Yankees ganaron y X es la proposición de que mi periódico reporta que ganaron.) En general, las correlaciones probabilísticas (condicionales e incondicionales) pueden proporcionar evidencia útil sobre las relaciones causales entre un conjunto de variables. Algunos filósofos incluso han definido la causalidad en términos de relaciones probabilísticas. Para más sobre todo esto, véase Hitchcock (2012).

Un punto final sobre las credibilidades condicionales. Antes (página 11) mencioné la teoría de que una credibilidad de x en P es simplemente la creencia plena de que la probabilidad de P es x . Ahí señalé varios problemas para esa teoría; ahora podemos añadir que a la teoría parece faltarle una buena manera de entender la credibilidad condicional. Una credibilidad condicional $c(P|Q)$ de x no puede leerse como una creencia cualitativa en la proposición “Si Q , entonces la probabilidad de P es x ”, ni tampoco puede leerse como la creencia de que “La probabilidad de ‘Si Q , entonces P ’ es x ”. Esto se estableció por medio de una serie de resultados de trivia-

13 Esta definición se generaliza al caso en el que Z es una variable aleatoria capaz de tomar diversos valores z_i . El filtrado de la correlación espuria ocurre cuando X y Y están incondicionalmente correlacionados, pero se vuelven independientes condicionales sobre cada proposición de la forma $Z = z_i$.

lidad por Lewis (1976).¹⁴ Por ejemplo, el trabajo de Lewis muestra que si se asume que $c(P|Q) = x$ sólo en caso de que $p(P \rightarrow Q) = x$ para alguna noción apropiada de probabilidad p y algún condicional indicativo $\rightarrow$, entonces se sigue que ¡cualquier proposición es probabilísticamente independiente de cualquier otra! Esto es obviamente absurdo. Una credibilidad condicional no es simplemente una credibilidad –o una creencia– sobre un condicional.

Actualizar por Condicionalización

Las restricciones racionales sobre la credibilidad enlistadas hasta ahora han sido sincrónicas –cuando relacionan múltiples credibilidades, todas las credibilidades relacionadas se tienen al mismo tiempo–. La literatura sobre grados de creencia también ha propuesto varias restricciones diacrónicas, que gobiernan las relaciones entre credibilidades asignadas en distintos momentos.

Supóngase que tenemos dos momentos, t_i y t_j , siendo el último posterior al primero. Sean c_i y c_j las funciones de credibilidad de la agente en estos dos momentos. La restricción credal diacrónica más tradicional, bien establecida y mejor conocida es Condiciona-

CONDICIONALIZACIÓN

Si $E \in \mathcal{L}$ representa todo lo que la agente aprende entre t_i y t_j , entonces para cualquier $X \in \mathcal{L}$, $c_j(X) = c_i(X|E)$.

La idea intuitiva de Condicionaización es simple. Supóngase que en t_i no sabes si E es verdadera. Te pido que hipotéticamente supongas E (que temporalmente lo añadas a tu repertorio de suposiciones sobre cómo es el mundo), luego te pregunto por tu credibilidad condicional en X dada esta suposición. Ofreces algún número. Entonces, entre t_i y t_j , descubres que E de hecho es verdadera (y no aprendes ninguna otra cosa). Si ahora te pregunto en t_j tu credibilidad incondicional en X , parece que deberías ofrecer el mismo número que reportaste como credibilidad condicional antes. Después de todo, el conjunto de condiciones del mundo real frente a las cuales evalúas X es el mismo en ambos momentos; sólo

14 Para el estado reciente de la cuestión en esta área, véase Hájek (2011) y Fitelson (2015).

que en t_i estabas suponiendo E como un hecho sobre el mundo mientras que en t_j sabes que E es verdadera.

La condicionalización se integra bien a nuestras otras restricciones credales. Por ejemplo, si c_i satisface los axiomas de Kolmogorov y $c_i(E) > 0$, entonces condicionalizar produce una distribución c_j que también satisface los axiomas. Así que, si una agente comienza con una distribución de probabilidad y actualiza reiteradamente al condicionalizar, está garantizado que continuará respetando el probabilismo. Los axiomas de probabilidad y la Fórmula de Proporción también hacen que actualizar por Condicionalización sea acumulativo y conmutativo. Si condicionalizas sucesivamente sobre E y luego E' , esto produce el mismo resultado que condicionalizar sólo una vez sobre $E \& E'$, lo cual significa que también produce el mismo resultado que condicionalizar sobre E' seguido de E .

Para una agente que condicionaliza, las credibilidades en el momento actual interactúan de una manera interesante con las predicciones sobre credibilidades futuras. Supóngase que una agente está segura en t_i de que sus credibilidades en t_j se formarán al condicionalizar sobre una proposición que aprenderá a partir de alguna partición finita particular (quizá realizará un experimento entre t_i y t_j , y las proposiciones en la partición representan todos sus posibles resultados). Asumiendo que ella cumple con unas cuantas condiciones plausibles adicionales, tal agente satisfará el Principio de Reflexión.

PRINCIPIO DE REFLEXIÓN

Para cualquier $X \in \mathcal{L}$, $c_i(X | c_j(X) = r) = r$.

Este principio, introducido por van Fraassen (1984), establece que la credibilidad incondicional de la agente en X en t_i es igual a su expectativa en t_i de su credibilidad incondicional en X en t_j .¹⁵ Nota que aunque una c_j aparece en la expresión del lado derecho, el principio gobierna interacciones credales *sincrónicas*: relaciona las credibilidades c_i de la agente en X con sus credibilidades c_i sobre sus credibilidades futuras en X . Dadas (de nuevo) unas cuantas condiciones adicionales, el Principio de Reflexión

15 Para ver por qué, vuelve a nuestra formulación de la Ley de la Probabilidad Total en la página 21, y haz que cada T_i ahí afirme que la credibilidad incondicional de la agente en X asumirá en t_j algún valor real específico r .

puede derivarse a partir de los axiomas de Kolmogorov, la Fórmula de Proporción y la certeza de la agente de que actualizará al condicionalizar sobre algún miembro de una partición específica. No obstante, van Fraassen argumenta en la dirección opuesta: proporciona motivación independiente a favor el Principio de Reflexión, luego considera a la Condicionalización como una consecuencia derivable. Para más argumentos en ambas direcciones, y las condiciones adicionales requeridas, véase Weisberg (2007) y Briggs (2009).

Cuando una agente reiteradamente actualiza por Condicionalización, a menudo se encuentra a sí misma calculando el valor de $c(X|E)$. Este cálculo puede optimizarse por medio de un famoso teorema.

TEOREMA DE BAYES

Para cualesquiera $X, E \in \mathcal{L}$ con valores c distintos de cero,

$$c(X|E) = \frac{c(E|X) \cdot c(X)}{c(E)}.$$

El teorema de Bayes ha resultado tan central a la aplicación de Condicionalización que a los teóricos que trabajan con grados de creencia a menudo se les llama “bayesianos” (o “bayesianos subjetivos”, o “epistemólogos bayesianos”). En un momento describiré por qué el Teorema de Bayes es tan útil. Pero primero, vale la pena notar que el Teorema de Bayes es en efecto un *teorema*, fácilmente derivable a partir de los axiomas de Kolmogorov y la Fórmula de Proporción. El bayesianismo¹⁶ ha generado mucha controversia, especialmente entre los estadísticos. Pero la afirmación controversial en el bayesianismo no es que el Teorema de Bayes es verdadero. Todos concuerdan en que el Teorema se sigue a partir de los axiomas de Kolmogorov, y en que, *si* una agente ha de generar nuevas credibilidades con el tiempo al condicionalizar, *entonces* el teorema proporciona una herramienta versátil para calcular credibilidades post-actualización a partir de credibilidades pre-actualización. La controversia está en si las agentes realmente deberían actualizar sus credibilidades mediante Condicionalización y en si

16 Tradicionalmente se le atribuye el teorema al reverendo Thomas Bayes. Aunque Bayes nunca publicó el teorema, Richard Price lo encontró en sus notas y lo publicó tras la muerte de Bayes en 1761. Tiempo después, Pierre-Simon Laplace redescubrió el teorema de manera independiente y fue responsable de mucha de su popularización inicial.

la mejor manera de entender la inferencia científica es como una serie de condicionalizaciones.

Dejando esta controversia de lado, ¿por qué el análisis particular de $c(X|E)$ en el Teorema de Bayes es tan útil? Considera un contexto científico, en el cual una teórica tiene una partición finita de hipótesis $H_1, \dots, H_n$ sobre qué ocurre con algún fenómeno. La teórica planea efectuar un experimento que ella espera que le ayude a discriminar entre las hipótesis. En el momento t_p antes de que efectúe el experimento, la teórica tiene un conjunto de credibilidades incondicionales c_i a las que llamaremos sus ‘credibilidades *iniciales*’ (*priors*). La teórica efectúa el experimento entre t_i y t_p y supongamos que se representa a las observaciones que hace por medio de la proposición E . Dada esta nueva evidencia, Condicionización le ayuda a calcular sus credibilidades en t_p a las que llamaremos sus ‘credibilidades *posteriores*’ (*posteriors*).

Supóngase que nos interesa la convicción de la teórica en una hipótesis particular H_m después de que se obtengan los resultados experimentales. Al aplicar Condicionización, el Teorema de Bayes y luego la Ley de la Probabilidad Total al denominador del Teorema de Bayes, derivamos:

$$c_j(H_m) = \frac{c_i(E|H_m) \cdot c_i(H_m)}{c_i(E|H_1) \cdot c_i(H_1) + \dots + c_i(E|H_n) \cdot c_i(H_n)}. \quad (12)$$

Considera uno a uno los componentes de la fracción en el lado derecho. Primero, tenemos varias expresiones de la forma $c_i(H_x)$. Éstas son las credibilidades iniciales de la teórica en varias hipótesis. Presumiblemente, al realizar el experimento ella tiene algunos niveles de convicción incondicional en las hipótesis que está considerando; éstos proporcionan las credibilidades iniciales en cuestión. Luego tenemos expresiones de la forma $c_i(E|H_x)$. A la credibilidad condicional de una agente en un resultado experimental E dada alguna hipótesis H_x se le llama su *verosimilitud* (*likelihood*)* para esa evidencia bajo esa hipótesis. Una hipótesis científica bien definida debería hacer una predicción de cómo resultará el

* Nota del traductor: Sigo la convención de emplear ‘*verosimilitud*’ para traducir al castellano la expresión ‘*likelihood*’ en este sentido técnico. Michael Titelbaum enfatiza que éste es un concepto distinto al de ‘proximidad a la verdad’, que en inglés es ‘*verisimilitude*’. Sobre esta cuestión terminológica consulté a Graham Oppie, quien claramente distingue ambos conceptos en su obra.

experimento de la teórica, o al menos debería asignar probabilidades a los varios posibles resultados. Éstas informan las verosimilitudes de la teórica para los varios resultados experimentales (como E) bajo las varias hipótesis que ella considera. Así, el Teorema de Bayes permite a la teórica formarse una opinión posterior sobre cada hipótesis H_m que ella considera, sobre la base de la evidencia que ha recibido, sus credibilidades iniciales incondicionales en las hipótesis y sus verosimilitudes en t_i —elementos que, presumiblemente, están todos fácilmente disponibles—.

Condicionaización de Jeffrey

A los estadísticos y a los filósofos de la ciencia a veces les preocupa que Condicionaización permita que el veredicto final de una científica sobre una hipótesis sea influenciado por sus credibilidades iniciales en esa hipótesis—su grado de creencia personal en la hipótesis antes de que aparezca cualquier evidencia—. A los epistemólogos les preocupa la concepción de la evidencia de Condicionaización. Parece que para que Condicionaización funcione, debe ser posible identificar alguna proposición E que represente *todo* lo que la agente aprende entre t_i y t_j . Además, la agente debe obtener *certeza* sobre E entre t_i y t_j , pues actualizar la credibilidad de la agente en E mismo usando Condicionaización produce que $c_j(E) = 1$. Finalmente, una vez que una agente tiene certeza de alguna proposición, actualizaciones subsecuentes por medio de Condicionaización mantendrán esa certeza por siempre.¹⁷

Por ende, Condicionaización parece incorporar una concepción del aprendizaje en la que lo que se aprende es explícitamente resumible en forma proposicional, se vuelve certeza y se retiene por siempre después. Para los epistemólogos, esto recuerda a aproximaciones fundacionistas a la evidencia abandonadas desde hace décadas. También viola el Principio de Regularidad, que juzga irracional que una agente le asigne certeza absoluta a una proposición empírica. (Después de todo, ¿qué evidencia podría hacer que estuvieras *completamente seguro* de que alguna proposición empírica es verdadera?).

17 Es fácil mostrar que si una agente condicionaliza sobre E entre t_i y t_j , tendrá $c_j(E) = 1$, y si luego condicionaliza sobre otra evidencia entre t_j y t_k , aun tendrá $c_k(E) = 1$ también.

Para atender estos problemas, Richard C. Jeffrey ofrece una regla de actualización que generaliza Condicionización para permitir experiencias de aprendizaje en las que no se obtengan certezas. Él introduce su regla usando el siguiente ejemplo.

La agente inspecciona un pedazo de tejido a la luz de una vela, y le da la impresión de que es verde, aunque admite que podría ser azul o incluso (aunque muy improbablemente) morado. Si V , A y M son las proposiciones de que el tejido es verde, azul y morado, respectivamente, entonces el resultado de la observación podría ser que, mientras originalmente sus grados de creencia en V , A y M eran .30, .30 y .40, sus grados de creencia en esas mismas proposiciones después de la observación son .70, .25 y .05 (Jeffrey, 1965, p. 154).

Al discutir el ejemplo, Jeffrey escribe:

Si hubiera una proposición E en la jerarquía de preferencias [de la agente] que describiera la cualidad precisa de su experiencia visual al ver el tejido, se diría que lo que la agente aprendió de la observación fue que E es verdadera... Pero no necesita haber tal proposición E en su jerarquía de preferencias; ni se requiere que ninguna de tales proposiciones sea expresable en castellano. La descripción ‘El tejido se veía verde o posiblemente azul o concebiblemente morado’, sería demasiado vaga para transmitir la cualidad precisa de la experiencia... Parece que lo mejor que podemos hacer es describir, no la cualidad de la experiencia visual en sí misma, sino más bien sus efectos sobre la observadora, al decir “Tras la observación, los grados de creencia de la agente en V , A y M eran .70, .25 y .05” (Jeffrey, 1965, p. 154-5).

Jeffrey propuso una regla de actualización que llamó “cinemática de la probabilidad”; hoy en día todos la llaman “Condicionización de Jeffrey”. La regla se aplica cuando la experiencia de una agente afecta sus credibilidades al alterar su distribución de grados de creencia a través de una partición finita particular en $\mathcal{L}$; cualesquiera otros cambios en sus credibilidades son causados por los cambios a esta partición. Si la partición de origen es $B_1, \dots, B_n$, entonces la regla de Jeffrey es como sigue.

CONDICIONALIZACIÓN DE JEFFREY

Para cualquier $A \in \mathcal{L}$, $c_i(A) = c_i(A|B_i) \cdot c_j(B_i) + \dots + c_i(A|B_n) \cdot c_j(B_n)$.

Jeffrey no pretendía descartar la posibilidad de que algún aprendizaje ocurra mediante la adquisición de certeza. Él sólo quería permitir la posibilidad de otros tipos de experiencias de aprendizaje también. Así que en el caso de que una de las B_m se vuelve certeza (y por ende todos los otros miembros de la partición obtienen credibilidad 0), Condicionación de Jeffrey se reduce a Condicionación tradicional.

Veamos cómo Condicionación de Jeffrey se aplica al ejemplo de Jeffrey del tejido a la luz de la vela. Supóngase que la agente se interesa en la proposición T , de que el pedazo de tejido seleccionado combinará con su sillón. Está segura de que cualquier morado combinará, está segura de que el verde no y tiene 50% de convicción en que un tejido azul combinará (que combine depende del tono específico de azul). Sea t_i el tiempo antes de que inspeccione el tejido a la luz de la vela. Al usar la Ley de la Probabilidad Total y las credibilidades incondicionales iniciales que Jeffrey proporciona, tenemos

$$\begin{aligned} c_i(T) &= c_i(T|V) \cdot c_i(V) + c_i(T|A) \cdot c_i(A) + c_i(T|M) \cdot c_i(M) \\ &= 0 \cdot .30 + .50 \cdot .30 + 1 \cdot .40 = 0.55. \end{aligned} \quad (13)$$

Jeffrey también proporciona las credibilidades incondicionales de la agente en V , A y M en t_j , luego de la inspección. Con estos valores, Condicionación de Jeffrey arroja

$$\begin{aligned} c_j(T) &= c_j(T|V) \cdot c_j(V) + c_j(T|A) \cdot c_j(A) + c_j(T|M) \cdot c_j(M) \\ &= 0 \cdot .70 + .50 \cdot .25 + 1 \cdot .05 = 0.175. \end{aligned} \quad (14)$$

El vistazo a la luz de la vela incrementa la convicción de la agente en que el tejido es verde y disminuye su convicción en que el tejido es morado, de modo que Jeffrey recomienda que disminuya la credibilidad posterior en que el tejido combine.

Nota cómo se produce este cambio en la credibilidad. La experiencia visual de la agente cambia sus credibilidades alterando directamente su distribución a través de la partición de color del

tejido. Cualesquiera cambios a otras proposiciones en el lenguaje de la agente (tales como T) son efectos ulteriores consecuencia de esta alteración directa. Aun así, las dependencias directas entre estas proposiciones subsecuentes y las proposiciones sobre color permanecen inalteradas: cambiar las opiniones de la agente sobre el color del tejido no cambia cuánta convicción tiene en que los colores particulares combinarán con el sillón. Ésta es la razón por la que las mismas credibilidades condicionales aparecen tanto en los cálculos de $c_i(T)$ como en los de $c_j(T)$.

Contra el trasfondo de los axiomas de Kolmogorov y la Fórmula de Proporción, Condicionalización de Jeffrey es equivalente a la siguiente condición.

RIGIDEZ

Para cualquier $A \in \mathcal{L}$ y cualquier B_m , $c_j(A | B_m) = c_i(A | B_m)$.

En una Condicionalización de Jeffrey, la experiencia altera las credibilidades de una agente a través de la partición B . Las credibilidades de la agente en otras proposiciones condicionales sobre las B_m s no cambian. Así que la agente establece sus credibilidades posteriores al adoptar credibilidades incondicionales en las B_m s a partir de la experiencia, que se copian sobre sus credibilidades condicionales antiguas, y luego aplicando la Ley de la Probabilidad Total para calcular sus credibilidades incondicionales en proposiciones no- B .

Requisitos racionales adicionales

Hasta ahora hemos visto varias presuntas restricciones racionales sobre la credibilidad: los axiomas de la probabilidad, la Fórmula de Proporción, el Principio de Reflexión, Regularidad y las reglas diacrónicas de Condicionalización y Condicionalización de Jeffrey. Aun así, hay un número infinito de distribuciones de credibilidad (y de secuencias de distribuciones de credibilidad a través del tiempo) compatibles con estas restricciones. ¿Son racionalmente permisibles todas esas distribuciones? Algunas de ellas parecen bastante extrañas y poco intuitivas —por ejemplo, algunas asignan muy alta credibilidad a escenarios escépticos; algunas llevarán a las agentes a razonar contra-inductivamente—.

A una posición extrema sobre la fuerza de las restricciones racionales suele llamársele “Bayesianismo objetivo”. Esta posición

adopta la Tesis de la unicidad (Feldman, 2007; White, 2005), la cual afirma que, dado cualquier cuerpo de evidencia, hay exactamente una distribución de credibilidad racionalmente permitida a cualquier agente con ese cuerpo de evidencia total. En el otro extremo, los que podríamos llamar “Bayesianos subjetivos extremos” sostienen que cualquier distribución de credibilidad es racionalmente permisible. Entre ellos están los “Bayesianos subjetivos moderados”, quienes sostienen que hay algunas restricciones racionales más allá de las que hemos descrito, pero no suficientes para generar una única distribución permisible en todos los casos.

¿Cuáles podrían ser estas restricciones racionales adicionales? Una restricción que podría estrechar considerablemente el campo de lo que es racionalmente permisible es el

PRINCIPIO DE INDIFERENCIA

Si una agente no tiene evidencia a favor de ninguna posibilidad en una partición por encima de cualquier otra, entonces debería asignar igual credibilidad a cada elemento de la partición.¹⁸

La objeción tradicional a este principio es que parece ofrecer recomendaciones en conflicto cuando volvemos a particionar el mismo espacio de posibilidades. Siguiendo a van Fraassen (1989), supón que te digo que un cubo ha sido producido en una fábrica y que la longitud de su lado está entre 0 y 1 metros. Dada la escasez de evidencia adicional, si te pregunto qué tan convencido estás de que la longitud del lado es menor que 0.5 metros, el Principio de Indiferencia parece exigir una credibilidad de $1/2$. Pero si ahora te pregunto qué tan convencido estás de que el volumen (que debe ser de entre 0 y 1 metros cúbicos) es menor que 0.5 metros cúbicos, el Principio de Indiferencia también parece requerir una credibilidad de $1/2$. Puesto que una longitud por lado de 0.5 metros corresponde a un volumen de 0.125 metros cúbicos, la única manera de asignar estas credibilidades consistentemente con los axiomas de la probabilidad es estar absolutamente seguro de que el volumen en metros cúbicos no está entre 0.125 y 0.5!¹⁹

18 La idea básica aquí se remonta al menos a Laplace (1814/1995), quien la vio como una aplicación de lo que Bernoulli (1713) llamaba el “principio de razón insuficiente”.

19 Un primo técnicamente más sofisticado del Principio de Indiferencia es el Principio de Máxima Entropía de Jaynes (1957a, 1957b). Este principio se

Otra familia de presuntas restricciones racionales tiene un miembro que ya hemos visto. El Principio de Reflexión nos indica que debemos fijar nuestras credibilidades incondicionales actuales en una proposición en un valor igual al que estamos seguros que tendrán en el futuro –o, si no estamos seguros de nuestras credibilidades futuras, igual a nuestra expectativa de cuáles tendrán–. Este principio nos indica que seamos deferentes ante las opiniones de nuestro yo futuro como si se tratara de alguna clase de experto. Pero por supuesto que hay otros expertos en el mundo, como aquellos de nuestros contemporáneos que creemos que tienen mejor juicio o mejor información que nosotros. Siguiendo el ejemplo del Principio de Reflexión, Elga (2007) sugiere que si c_e es la distribución de credibilidad de una agente a quien consideramos una experta, entonces para cualquier $X \in \mathcal{L}$ (o al menos cualquier X en el área de competencia de la experta) deberíamos asignar

$$c(X) \mid c_e(X) = r) = r. \quad (15)$$

Pensando de manera más metafórica, una distribución “experta” digna de nuestra deferencia no necesita siquiera ser una agente. Podría ser racional alinear nuestras credibilidades con ciertos valores numéricos objetivos en el universo. Esto nos lleva al tema de los principios de inferencia directa.

Principios de inferencia directa

En la página 10 brevemente se mencionaron las interpretaciones de la probabilidad –propuestas para el significado de locuciones sobre “probabilidad”–. Por ejemplo, la interpretación clásica, que data al menos de Laplace (1814), definía la probabilidad como el número de resultados favorables de un proceso dividido por el número total de posibles resultados. Después, la teoría de la probabilidad como frecuencia (asociada más estrechamente con von Mises, 1928), leía probabilidad como la frecuencia con la

aplica de manera más natural a particiones infinitas y se adapta bien a varias formas de evidencia. Aun así sucumbe a los problemas de varianza de partición y también está en conflicto con actualizar por condicionalización en casos particulares. Véase Seidenfeld (1986).

que un resultado ocurriría si un proceso particular se repitiera muchas veces.²⁰

Mi tarea aquí no es evaluar estas nociones de probabilidad como propuestas en la teoría del significado o en la teoría de “probabilidad”. Más bien, quiero preguntar qué tienen que ver estas nociones con la credibilidad racional. Muchos bayesianos han suscrito principios de inferencia directa: principios que llevan a la agente a partir de información sobre alguna noción de probabilidad a credibilidades específicas en sucesos específicos. Por ejemplo, podría ser que, si yo estoy seguro de que un tipo particular de configuración experimental produce un tipo particular de resultado con la frecuencia x , entonces cuando vaya a efectuarse un experimento de ese tipo, yo debería tener la credibilidad x de que producirá un resultado de ese tipo. Éste sería un principio de inferencia directa de hechos sobre frecuencia a credibilidades en resultados.

Los principios de frecuencia-a-credibilidad enfrentan dificultades notables, incluso cuando se esbozan de manera tan burda como lo acabo de hacer. Como muestra, un único suceso (voy a mi oficina mañana) puede clasificarse como el resultado de varios tipos de experimentos (elegir si iré en un día de verano, elegir si iré un martes, etc.), los cuales pueden producir distintas frecuencias y, por ende, diferentes recomendaciones credales. (Ésta es una versión del “problema de la clase de referencia”²¹). También, si intentáramos usar este principio como una estrategia general de fijación-de-credibilidades, tendríamos problemas con experimentos que parecen ser irrepetibles. Antes de que se encendiera el Gran Colisionador de Hadrones, los diarios reportaron profusamente los grados de creencia de los físicos de que hacerlo destruiría la Tierra. Es difícil alinear tales credibilidades con la frecuencia con la que encender el colisionador causaría destrucción global; en caso de tal destrucción, encenderlo sólo ocurre una vez.

Por ende, podría ser preferible vincular la credibilidad racional con el “azar objetivo”. Como noción de probabilidad, el

20 La sección anterior introdujo un uso de la terminología “Bayesianismo objetivo/subjetivo”. Ese uso debería distinguirse claramente de otro que a menudo aparece en la literatura sobre interpretaciones de la probabilidad. En esa literatura, “Bayesianismo subjetivo” describe la posición de que, en el habla cotidiana, “probabilidad” siempre se refiere a o expresa credibilidades subjetivas. El “Bayesianismo objetivo”, por otra parte, sostienen que el habla sobre probabilidad refiere a algo más allá del sujeto, como las frecuencias o el azar.

21 Véase Hájek (2007) para muchas más versiones.

azar es objetivo en el sentido de que su valor está determinado por la constitución física de un aparato experimental. El azar también puede aplicarse a sucesos que sólo ocurren una vez. Un principio de frecuencia-a-credibilidad recomienda una credibilidad de $1/6$ en que un lanzamiento de dado no amañado resultará en 3, sobre la base de que repetir el lanzamiento producirá 3 en un sexto de las ocasiones. El teórico del azar objetivo recomienda $1/6$ sobre la base de que un dado no amañado está constituido físicamente de una manera particular (con igual peso en cada lado, etc.). Esto seguiría siendo verdad incluso si el dado nunca hubiese sido lanzado antes y fuese a ser destruido con certeza tras el lanzamiento en cuestión.

El principio de inferencia directa más famoso que vincula credibilidad con azar es el Principio Principal (*Principal Principle*) de Lewis (1980). *Muy* burdamente, y pasando por alto muchísimos detalles,²² el Principio Principal le indica a una agente que establezca

$$c(A \mid Ch(A) = x) = x. \quad (16)$$

a menos que posea evidencia inadmisible relevante para A . Aquí $Ch(A) = x$ es la proposición de que el azar objetivo de A es x . Así que —dejando de lado la cuestión de la evidencia inadmisible por un momento—, si la agente está segura de que, digamos, el resultado por azar de obtener 3 al lanzar un dado particular es $1/6$, el Principio Principal establecerá su credibilidad en 3 a $1/6$. Si, por otra parte, la agente sabe que el dado está amañado, pero divide equitativamente su credibilidad entre que resulte 3 en $1/10$ por azar y en $1/5$ por azar, la Ley de la Probabilidad Total se combinará con el Principio Principal para producir:

$$c(3) = c(Ch(3) = 1/10) \cdot c(3 \mid Ch(3) = 1/10) + c(Ch(3) = 1/5) \cdot c(3 \mid Ch(3) = 1/5) = 1/2 \cdot 1/10 + 1/2 \cdot 1/5 = 0.15. \quad (17)$$

Con otras palabras, su credibilidad en que el dado resultará en 3 es su expectativa del azar objetivo de obtener un 3. Por ende, podemos pensar en el Principio Principal como un principio de deferencia en el que el experto es el azar objetivo.

La innovación clave del Principio Principal de Lewis es su tratamiento de la evidencia que la agente considera relevante para el resultado de un suceso azaroso. Lewis divide tal evidencia en dos clases: la evidencia admisible es aquella que la agente con-

22

Véase Meacham (2010) para algunos de esos detalles.

sidera relevante al resultado debido a que afecta su opinión sobre el azar objetivo del suceso. Por ejemplo, la información sobre el peso del dado es admisible con respecto al resultado del lanzamiento —afecta cómo piensa la agente que resultará el lanzamiento al afectar lo que la agente piensa que es el azar de un 3—. La evidencia inadmisble afecta la opinión de la agente de alguna otra manera. Por ejemplo, si un soldado confederado le dice el resultado de su lanzamiento, esto afecta la opinión de la agente sobre si resultó 3, pero no al hacerle pensar que el azar objetivo de 3 era distinto. La innovación de Lewis fue que los hechos de azar sobre un resultado filtran correlaciones espurias de información admisible relevante para un resultado. Así que, si E es admisible, el Principio Principal también nos da:

$$c(A | Ch(A) = x \ \& \ E) = c(A | Ch(A) = x) = x. \quad (18)$$

La evidencia admisible se relaciona con el azar objetivo como la causa distante de un suceso se relaciona con la causa próxima.

Aditividad Contable

Hasta este punto los ejemplos que hemos considerado típicamente involucran sólo cantidades finitas de posibilidades. Pero ¿qué tal si una agente considera una partición de infinitos posibles resultados y distribuye equitativamente su credibilidad entre ellos? ¿Cómo puede modelarse esto en nuestra epistemología bayesiana?

Para poner un ejemplo concreto, supongamos que se ha elegido un número entero positivo mediante un proceso, y nuestra agente quiere asignar igual credibilidad a cada entero de ser seleccionado. Presumiblemente, eso debería ser posible. Pero, ¿qué valor numérico podría tomar esa credibilidad? Es fácil mostrar que los axiomas de la probabilidad impiden que sea un número real positivo. Pues supóngase que la agente asigna

$$r = c(1) = c(2) = c(3) = \dots \quad (19)$$

(Donde $c(1)$ es su credibilidad en que 1 fue elegido).²³ Para cualquier número real positivo r , habrá un entero positivo n tal que $r > 1/n$. Ahora considera la credibilidad de la agente de que el entero seleccionado esté entre 1 y n (incluyéndolos). Si vuelves a ver la lista de restricciones intuitivas que se siguen de los axiomas de Kolmogorov (páginas 15 a 19), el último principio en la lista con viñetas nos dará

$$c(1 \vee 2 \vee \dots \vee n) = c(1) + c(2) + \dots + c(n) = r \cdot n > 1, \quad (20)$$

lo cual viola los axiomas.

¿Cuáles son las opciones disponibles? Una sugerencia popular es que, cuando una agente asigna igual convicción a las infinitas posibilidades, representemos ese nivel de convicción como una credibilidad de 0. Así, diríamos que

$$c(1) = c(2) = \dots = 0.$$

Usar la credibilidad 0 de esta manera introduce unos cuantos problemas. Primero, hasta este punto hemos concebido a la credibilidad 1 como representado la certeza en una proposición y la credibilidad 0 como certeza en que la proposición es falsa. Ahora tendremos que permitir a una agente asignar $c(P) = 0$, incluso si la agente admite que P podría ser verdadera, y $c(\neg P) = 1$, incluso si la agente no está segura de que P es falsa. Y tendremos que enunciar el Principio de Regularidad con cuidado: puede que aun queramos prohibir a las agentes asignar certeza a proposiciones empíricas, sin censurar credibilidades de 1 y 0 a tales proposiciones.

Segundo, la Fórmula de Proporción que ofrecimos sólo relaciona la credibilidad condicional $c(X|I)$ con credibilidades incondicionales cuando $c(I) > 0$. Necesitaremos expandir este principio para manejar casos en los que $c(I) = 0$ y aun así la agente no descarta I . Por ejemplo, nuestra agente que asigna igual credibilidad a la selección de cada entero positivo podría asignarle $c(2|2 \vee 4) = 1/2$, aunque $c(2 \vee 4) = c(2) + c(4) = 0$.²⁴

23 Nota que ahora estamos tratando con un lenguaje que contiene un número infinito de fórmulas atómicas. Aunque éste es un cambio con respecto a nuestro escenario anterior, no es difícil de manejar y es bastante común en modelos formales.

24 Una manera de manejar esta situación es considerar a las credibilidades condicionales como básicas. Ver la nota 10 en la página 20 para más información.

Tercero y más importante, queremos una manera de sumar credibilidades a través de disyunciones infinitas. Aditividad Finita sólo cubre disyunciones con una cantidad finita de disyuntos –¿qué tal si queremos calcular la credibilidad de nuestra agente en que el número entero seleccionado es par?–. Una extensión natural de Aditividad Finita es la siguiente.

ADITIVIDAD CONTABLE

Para cualquier partición contable $\{Q_1, Q_2, Q_3, \dots\} \subset \mathcal{L}$,

$$c(Q_1 \vee Q_2 \vee Q_3 \vee \dots) = c(Q_1) + c(Q_2) + c(Q_3) + \dots$$

Aditividad Contable no sólo es natural; también nos permite establecer una restricción muy importante sobre las credibilidades.

CONGLOMERABILIDAD

Para cualquier proposición $P \in \mathcal{L}$ y partición $\{Q_1, Q_2, Q_3, \dots\} \subset \mathcal{L}$, $c(P)$ no es mayor que la $c(P|Q_i)$ más grande ni menor que la $c(P|Q_i)$ más pequeña.

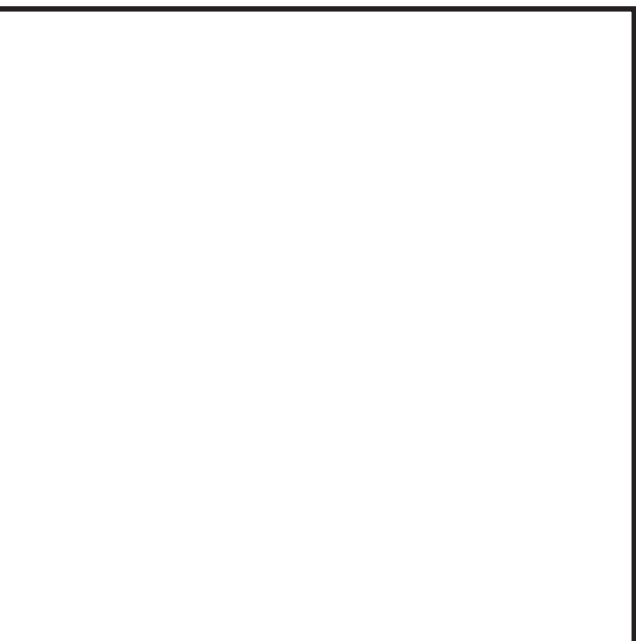
Dada Conglomerabilidad, las $c(P|Q_i)$ establecen límites superiores e inferiores sobre el valor de $c(P)$. Esto tiene sentido si piensas en $c(P)$ como la media ponderada de las credibilidades que la agente asignaría a P condicional sobre todas las diferentes Q_i posibles. Y es especialmente importante cuando la agente tiene una partición $\{E_1, E_2, E_3, \dots\}$ de posibles nuevas piezas de evidencia que podría recibir antes de la siguiente actualización. Asumiendo que planea actualizar mediante Condicionización, ella sabe que su credibilidad futura en P será una de sus $c(P|E_i)$; luego Reflexión exige que satisfaga Conglomerabilidad.²⁵

El paquete Conglomerabilidad/Aditividad Contable es atractivo. Pero es inconsistente con asignarle una credibilidad de 0 a cada entero positivo en nuestro ejemplo. La razón es simple:

25 Nota que mi enunciación de Conglomerabilidad no especifica la cardinalidad de la partición Q_i . Para particiones finitas, Conglomerabilidad puede demostrarse a partir de los axiomas de probabilidad estándar. Adoptar Aditividad Contable extiende Conglomerabilidad a particiones contables. Para una agente que considera disyunciones más grandes que éstas, Seidenfeld, Schervish y Kadane (2017) muestran que en cada cardinalidad requerimos el principio relevante de Aditividad para asegurar Conglomerabilidad en particiones de ese tamaño.

dada Aditividad Contable, la credibilidad de la agente en que algún entero positivo será elegido es la suma de sus credibilidades en cada entero individual. Pero el primer valor debería ser 1, mientras que los últimos valores individuales son cada uno de 0. Así que los defensores de Aditividad Contable han sugerido en su lugar que en esta situación la agente asigne un valor infinitesimal a cada entero que es seleccionado. Los infinitesimales son una extensión del conjunto de los números reales, definidos como mayores que 0 pero menores que cualquier número real. Así, no caen presa del problema de nuestra Ecuación 20. Al mismo tiempo, sumar una cantidad infinita de infinitesimales puede producir un número real, así que podemos mantener tanto Aditividad Contable como una credibilidad de 1 en que algún entero será elegido.

Aun así, los infinitesimales introducen sus propias dificultades; para algunas, y muchos de los detalles matemáticos, véase Hájek (2003, Sec. 5), Williamson (2007), Easwaran (2014) y Wenmackers (2019).



Aplicaciones de la credibilidad

He presentado el estudio bayesiano de la credibilidad como el estudio de un tipo de actitudes doxásticas y de lo que se requiere para que tales actitudes sean racionales. Este estudio es valioso por derecho propio, como una contribución a la epistemología y la filosofía de la mente. Pero históricamente también se ha emprendido para mejorar nuestra comprensión de otros temas; discutiremos algunos de ellos en esta sección.

La teoría de la confirmación

Un epistemólogo o un filósofo de la ciencia bayesiano estudia la justificación y el apoyo evidencial al pensar acerca de la “confirmación”. El tipo de confirmación estudiado es usualmente incremental, más que absoluta (*all-things-considered*); cuando decimos que “la evidencia E confirma la hipótesis H ”, queremos decir que E proporciona al menos algún apoyo evidencial positivo a favor de H , no que resuelve la discusión sobre H ni siquiera que lleva a H por encima de algún umbral crucial.²⁶ Para un bayesiano, la confirmación es siempre también relativa a una distribución de probabilidad, y a un corpus de proposiciones de trasfondo. Por lo general, la distribución de probabilidad será la función de credibilidad de algún agente y el corpus de trasfondo será la evidencia total que informa esa función de credibilidad. (Bajo un régimen de Condicionización, el corpus se representa formalmente por me-

26 Esto contrasta con la manera en la que a veces se usa “confirmada” en castellano, como cuando hablamos de que una nominación, o incluso una reservación, ha sido confirmada.

dio del conjunto de todas las proposiciones X tales que $c(X)=1$.²⁷ Así que tomamos a una agente específica en un momento dado y preguntamos si E confirma H para ella, relativo a sus credibilidades y al corpus de trasfondo en ese momento.

Dejando que K represente un corpus de trasfondo, y que c_k represente una distribución de probabilidad informada por ese corpus, la teoría bayesiana de la confirmación postula que

E confirma H relativo a c_k sólo en caso de que $c_k(H|E) > c_k(H)$.

La confirmación bayesiana es simplemente relevancia probabilística positiva relativa a c_k . (De manera análoga, la socavación [*disconfirmation*] se define habitualmente como relevancia negativa relativa a c_k).

Aunque bastante simple, esta teoría de la confirmación resulta ser sorprendentemente sutil, poderosa y convincente. Para ilustrarlo –y fijar en la mente del lector la noción deseada de apoyo evidencial– supón que un dado sin amañar acaba de ser lanzado, y no sabes nada sobre el resultado. Quizá, de acuerdo con el Principio Principal, algún principio de frecuencia, o incluso el Principio de Indiferencia, asignas igual credibilidad a cada uno de los seis posibles resultados. Relativo a tu distribución de credibilidades y tu corpus de trasfondo, si recibieras evidencia de que el lanzamiento resultó en un número primo, esto confirmaría para ti que el lanzamiento dio como resultado un número impar. ¿Por qué? Debido a que, si satisfaces los axiomas de Kolmogorov y la Fórmula de Proporción, entonces asignas

$$2/3 = c(\text{impar} | \text{primo}) > c(\text{impar}) = 1/2. \quad (21)$$

Esto no significa que la evidencia de que es primo deba hacer que estés *seguro* de que el lanzamiento resultó en un número impar, ni siquiera que te justificaría al *creer* que el lanzamiento resultó en impar. Pero si actualizas por Condicionización, aprender que el lanzamiento resultó en primo haría que estuvieses al menos un poco más convencido de que el lanzamiento resultó en impar. De nuevo, la confirmación aquí es incremental.

27 Nota que a pesar de nuestra sugerencia en la página 37 de que a veces podría interpretarse de otra manera, he vuelto a tratar la credibilidad de 1 como representando certeza. Para simplificar la discusión, continuaré haciéndolo en lo que sigue.

Esta teoría bayesiana de la confirmación le da a la relación de confirmación algunas propiedades formales interesantes e intuitivas.²⁸

- Si $E \models E'$ y $H \models H'$, entonces E confirma H sólo en caso de E' confirme H' .
- E confirma H sólo en caso de que E socave $\neg H$.
- Si E y $K \models H$, pero $K \not\models H$, entonces E confirma H .
- Si H y $K \models E$, pero $K \not\models H$, entonces E confirma H .

La primera de estas propiedades asegura que las proposiciones lógicamente equivalentes se comportan igual dentro de la relación de confirmación. La segunda vincula a la confirmación con la socavación. La tercera y la cuarta de estas propiedades²⁹ especifican cómo la confirmación se vincula con la consecuencia lógica. La tercera propiedad nos dice que la consecuencia lógica es una forma de confirmación; si E tiene como consecuencia H en conjunción con K , mientras K no tiene a H como una de sus consecuencias, entonces E confirma H . En cuanto a la cuarta propiedad, captura la idea³⁰ de que una hipótesis que (en conjunto con el corpus de trasfondo) predice una observación evidencial es confirmada por esa observación.

Por otra parte, la teoría bayesiana distancia a la relación de confirmación de algunas propiedades que en ocasiones se le adscriben erróneamente. Aquí hay dos ejemplos.

- Si E confirma tanto H como H' , entonces el conjunto H, H', K es lógicamente consistente.
- Si X confirma T y T confirma Z , entonces X confirma Z .

Es importante rechazar la primera de estas propiedades debido a que hablamos de confirmación incremental. Por ejemplo, en la ilustración de Jeffrey en la cual una agente inspecciona un pedazo de tejido a la luz de una vela, su breve vistazo podría con-

28 En cada una de estas propiedades, las expresiones “ E confirma H ” y “ E socava H ” debería estar seguida de la frase “relativo a c_k ”. En lo que sigue simplificaré las locuciones al dejar implícita la relativización a c_k siempre que sea posible.

29 Ambas requieren una condición adicional de que el conjunto $\{E, K, H\}$ sea lógicamente consistente.

30 La cual es familiar a partir del hipotético-deductivismo (Crupi, 2020, Sec. 2).

firmar que el tejido es verde, aunque también confirme que es azul o incluso que es morado. (Quizá el vistazo socava que el tejido sea rojo y socava también que sea anaranjado). Esto es perfectamente razonable, a pesar del hecho de que verde, azul y morado son hipótesis inconsistentes sobre el color del tejido. De manera análoga, en contextos científicos, la misma observación puede confirmar teorías mutuamente excluyentes dentro de una partición, mientras que al mismo tiempo (quizá) descarta otras.

La segunda propiedad es la presunta transitividad de la confirmación. Éste es uno de los errores que más comúnmente se cometen acerca de la confirmación, el apoyo, la justificación y otras nociones afines.³¹ El mero hecho de que X confirme Y y Y confirme Z no significa que X confirme Z —incluso en el caso especial en el que Y tiene como consecuencia lógica Z —. Para ver por qué, imagina que se extrae una carta al azar de una baraja estándar. La información de que la carta es de espadas confirma (¡incrementalmente!) que la carta es la sota de espadas. Pero la información de que la carta es de espadas no confirma ni siquiera incrementalmente que la carta es una sota.

Otro error común es confundir lo que Carnap (1962) llamó concepciones de la confirmación como “firmeza” y como “incremento en firmeza”.³² La explicación bayesiana que hemos estado discutiendo es una concepción de incremento en firmeza. Una concepción de firmeza, por otra parte, dice que E confirma H relativo a c_k sólo en caso de que $c_k(H|E)$ sea alta (donde la altura necesaria puede ser influenciada por, digamos, parámetros contextuales). Entre muchos otros problemas, la concepción de firmeza erra al sostener que E confirma H en casos en los que $c_k(H|E)$ es alta simplemente debido a que la $c_k(H)$ inicial es alta. De hecho, una concepción de firmeza puede decir que ¡ E confirma H relativo a c_k incluso si $c_k(H|E)$ es menor que $c_k(H)$ (mientras $c_k(H|E)$ siga siendo alta)! La concepción bayesiana se enfoca en la relación entre E y H —cómo E alteraría la opinión de la agente sobre H — más que sólo dónde caería esa opinión si E fuese tomada en cuenta.

Podemos proporcionar más información sobre el efecto de E en la opinión de la agente sobre H al medir el grado de confir-

31 Corregir este error ha sido un tema en la literatura en epistemología sobre clausura epistémica. Véase, *e.g.*, Dretske (1970), Davies (1998) y Wright (2003).

32 Carnap estaba familiarizado con este error; al haberlo cometido en la primera edición de su *Logical Foundations of Probability*, en 1950.

mación incremental. La manera más simple de medir la confirmación es calcular $c_k(H|E) - c_k(H)$; esta medida simplemente pregunta qué tanto condicionalizar sobre E incrementaría la convicción de la agente en H . Aun así, como medida del impacto de E sobre H , esta simple diferencia tiene algunos inconvenientes. Por ejemplo, el grado en el que E puede confirmar H estará limitado por el valor de $c_k(H)$. Si, digamos, $c_k(H) = 0.99$, entonces incluso si E tiene como consecuencia lógica H , el máximo grado en el que puede confirmar H será 0.01. La teoría de la confirmación bayesiana tiene por ende una literatura considerable que propone y evalúa medidas alternativas de fuerza confirmatoria; véase Crupi (2020, Sec. 3.4) para un panorama reciente y referencias.

Un resultado de la literatura sobre medición de la confirmación es una nueva aproximación a la “resolución” de las paradojas tradicionales de la confirmación. Por ejemplo, usualmente pensamos que las generalizaciones universales son confirmadas por sus casos positivos (*positive instances*). La hipótesis de que todos los cuervos son negros típicamente es confirmada por la evidencia de que un cuervo particular es negro.³³ En símbolos $(\forall x)(Cx \supset Nx)$, es confirmada por $Ca \& Na$. Pero ahora supóngase que descubrimos un objeto que es no-negro y no-cuervo. La evidencia $\neg Na \& \neg Ca$ es un caso positivo de la generalización $(\forall x)(\neg Nx \supset \neg Cx)$, de modo que debería confirmar esa generalización. Pero esta última generalización es (por contraposición) lógicamente equivalente a la primera. Así que, por la primera propiedad de la confirmación que adopté arriba, $\neg Na \& \neg Ca$ debería confirmar que todos los cuervos son negros. Ésta es la famosa “Paradoja de los cuervos”, de Hempel (1945), la cual parece generar la absurda conclusión de que una hipótesis sobre el color de los cuervos puede ser confirmada por la observación de un zapato blanco.

Recientemente, varios teóricos bayesianos han concedido que quizá un zapato blanco de hecho confirma que todos los cuervos son negros —es sólo que observar un zapato blanco confirma esta hipótesis mucho *menos* de lo que lo haría la observación de un

33 Digo “típicamente”, pues es posible generar un corpus de trasfondo anormal contra el cual sería razonable que la observación de un cuervo negro socavara que todos los cuervos son negros (para ejemplos, véase Swinburne, 1971 y Rosenkrantz, 1977, Chap. 2.). La generación de la paradoja no depende de tales corpora anormales, así que los dejaremos de lado por el resto de la discusión.

cuervo negro³⁴. Fitelson y Hawthorne (2010), por ejemplo, especifican condiciones sobre c_k tales que, mientras se cumpla con estas condiciones, la evidencia de un cuervo negro confirmará la hipótesis mucho más robustamente que un no-cuervo no-negro, virtualmente en cualquier medida de confirmación en la literatura. Es altamente plausible que la mayoría de nosotros en el mundo real tengamos distribuciones de credibilidad que satisfagan las condiciones de Fitelson y Hawthorne, explicando así nuestras intuiciones sobre la asimetría de preferencia en este caso. Aproximaciones similares se han tomado al problema de la conjunción irrelevante (Hawthorne y Fitelson, 2004) y a la paradoja de verzul (Chihara, 1981; Eells, 1982), de Goodman (1955).

La teoría de la decisión

Daré sólo un breve bosquejo aquí (ver Thoma, 2019). En la teoría formal de la decisión, una agente es confrontada con un problema de decisión, representado por una partición de actos que ella podría realizar. Una vez que ella realice un acto, un resultado ocurrirá, y la agente valora en diferentes grados los distintos resultados. Estas valuaciones son representadas por medio de una función de utilidad, que le asigna utilidades en números reales a cada posible resultado. (La suposición clave acerca de las utilidades es que miden el valor *uniformemente* —la agente considera que cada unidad de utilidad añadida es tan valiosa como la siguiente—. No ocurre lo mismo con el dinero; tu primer dólar puede ser mucho más valioso que el billonésimo).

Así que, ¿qué es difícil acerca de eso? —¿no debería la agente sólo elegir el acto que la lleve al resultado más valioso?—. El problema es que la agente puede estar incierta sobre qué actos llevarán a qué resultados. Para ponerlo de otra manera, la agente puede no estar segura de en qué estado del mundo se encuentra y el resultado que se sigue de su decisión puede depender tanto del acto que ella elija como del estado del mundo restante. Por ejemplo, supón que intento decidir si ir a mi oficina mañana. Sé que si voy *puede* que sea silencioso y tranquilo ahí, en cuyo caso podré avanzar mucho en la escritura, lo cual es un resultado al que asigno un elevado valor. Por otra parte, puede que haya una construcción ruidosa afuera de la ventana de mi oficina, en cuyo caso perderé el

34 Aunque la idea se remonta hasta Hosiasson-Lindenbaum (1940).

tiempo en internet y no escribiré, un resultado al que le asigno muy poca utilidad. Puesto que no sé cuál será el estado de construcción en torno a mi edificio mañana, no es claro para mí cuál acto disponible (ir a la oficina, quedarme en casa) se correlaciona con cuál resultado, lo que complica mi decisión.

La solución estándar a este problema es que la agente asigne un valor esperado (*expected value*) a cada acto disponible. El valor esperado de una agente para un acto es su expectativa para la cantidad de utilidad que se acumulará si realiza el acto —calculado usando sus credibilidades de que varios estados del mundo ocurrirán—. Dada una decisión entre dos actos, una agente racional prefiere el acto al que le asigna un valor esperado más alto (y es indiferente en caso de empates). Así, podemos usar sus credibilidades y sus asignaciones de utilidad para desarrollar un orden de preferencias sobre los actos que están disponibles para ella en cualquier problema de decisión.

Por ejemplo, supón que le asigno una utilidad de 100 a un día tranquilo de escritura en mi oficina, pero una utilidad 0 a pasar el día ahí cuando hay obras de construcción. Si tengo 40% de convicción en que no habrá construcción mañana, mi utilidad esperada de ir a la oficina es:

$$\begin{aligned} UE(\text{ir a la oficina}) &= c(\text{no construcción}) \cdot u(\text{escribir en calma}) + c(\text{construcción}) \cdot u(\text{día perdido}) \\ &= 0.40 \cdot 100 + 0.60 \cdot 0 = 40, \end{aligned} \tag{22}$$

donde la función u designa cuánta utilidad asigno a un resultado específico. Dada esta utilidad esperada para ir a la oficina, debería preferir quedarme en casa sólo si espero que hacerlo me produzca una utilidad mayor que 40.

Podemos demostrar que, si una agente establece sus preferencias al maximizar la utilidad esperada, su orden de preferencias sobre actos satisfará varias condiciones intuitivas, comúnmente conocidas como los “axiomas de preferencias”. Por ejemplo, sus preferencias serán asimétricas (nunca prefiere tanto A sobre B como B sobre A) y transitivas (si prefiere A sobre B y B sobre C , entonces prefiere A sobre C).

Como dije, evitaré las muchas sutilezas de desarrollar una teoría de la decisión completa. Una preocupación crucial son casos en los que el acto de la agente puede estar correlacionado con el

estado del mundo. Los teóricos de la decisión evidencial (Jeffrey, 1965) responden al trabajar con la credibilidad de la agente en un estado *condicional* sobre que ella realice un acto particular, mientras los teóricos de la decisión causal (Gibbard y Harper, 1978; Lewis, 1981; Joyce, 1999; Weirich, 2012) consideran la credibilidad de la agente de que su acto *causará* que ocurra un estado particular. Otra preocupación es modelar a las agentes propensas a evitar el riesgo —como una agente que prefiere una ganancia garantizada con utilidad 1 por encima del lanzamiento de una moneda ordinaria en el que cara ofrece un premio con utilidad 3 (Allais, 1953; Buchak, 2013)—.

Hay, sin embargo, una noción más de la teoría de la decisión que requeriremos en lo que sigue: precio de apuesta justo. Considera una proposición P y un boleto de apuesta que garantiza a quien lo posee \$1 si P resulta ser verdadera. ¿Qué tanto vale para ti ese boleto de apuesta? Eso depende de cuán convencido estés de P . Si estás seguro de que P , ese boleto vale \$1 para ti. Si estás seguro de que P es falsa, el boleto no vale nada para ti. Entre esos extremos, entre más convencido estés de P , mayor valor le asignarás al boleto de apuesta.

Para ser más preciso, tu valor esperado en dólares del boleto de apuesta justo es $c(P) \cdot \$1$. Llamamos a esto tu *precio de apuesta justo* para esta apuesta sobre P . En general, si una apuesta paga $\$X$ dólares cuando P es verdadera, tu precio de apuesta justo para la apuesta es

$$c(P) \cdot \$X. \quad (23)$$

¿Qué significa decir que éste es tu precio de apuesta justo? Supón que alguien ofrece venderte un boleto de apuesta que paga sobre P . Tu precio de apuesta justo es aquel con el que esperarías salir tablas (sin ganar ni perder) en tal inversión. Asumiendo que valoras el dinero de manera lineal (de modo que cada centavo adicional te confiere la misma cantidad de utilidad adicional), la teoría de la decisión dice que deberías estar dispuesta a *comprar* el boleto de apuesta por cualquier cantidad *inferior* a tu precio de apuesta justo, e indiferente sobre comprarlo a exactamente tu precio de apuesta justo. Conversamente, si posees tal boleto, deberías estar dispuesta a *venderlo* por cualquier cantidad *por encima* de tu precio de apuesta justo.

Otras aplicaciones

Históricamente, la confirmación y la teoría de la decisión han sido las principales impulsoras del desarrollo del bayesianismo y las dos aplicaciones más comunes a las que se ha sometido a esta aproximación. Pero la teoría bayesiana de las credibilidades también se ha aplicado a muchos otros temas filosóficamente importantes. Aquí hay unos cuantos ejemplos.

- Se han usado probabilidades para medir cuándo son coherentes las proposiciones en un conjunto. Se ha evaluado luego al coherentismo sobre la justificación al preguntar si la coherencia entre proposiciones hace racional conferir mayor credibilidad en cada una de ellas. Véase Shogenji (1999), Bovens y Hartmann (2003), Huemer (2011) y Olsson (2017).
- Se ha debatido si el hecho de que una agente actualice por condicionalización incrementará su credibilidad en la hipótesis que mejor explique la evidencia observada. Van Fraassen (1989) sostiene que el bayesianismo es incompatible con la inferencia a la mejor explicación. Respuestas han sido ofrecidas por, *inter alia*, Okasha (2000), Lipton (2004), Weisberg (2009) y Henderson (2013).
- Elga (2007) argumenta que cuando una agente descubre que un par epistémico ha asignado distintas credibilidades que ella basada en la misma evidencia, esa agente debería aproximar sus credibilidades a las de su par. Mucho debate ha surgido en torno a si tal conciliacionismo es la respuesta racional al desacuerdo entre pares. Christensen (2009) presenta un útil panorama que desafortunadamente ya es obsoleto; Christensen y Lackey (2013) es una colección más reciente (¡aunque mucho se ha publicado sobre el tema desde entonces!).
- La controversia sobre el desacuerdo entre pares se entrecruza con cuestiones más amplias sobre la respuesta racional a evidencia de orden superior –evidencia concerniente a si uno ha respondido racionalmente a la propia evidencia–. Pueden encontrarse nuevos ensayos sobre la evidencia de orden superior y su conexión con el desacuerdo en Rasmussen y Steglich-Petersen (2019).

- El desacuerdo entre pares es también un aspecto de la epistemología social, la cual ha considerado por décadas cómo grupos e individuos deberían combinar las opiniones de múltiples expertos para formar una única perspectiva coherente. La literatura sobre la acumulación de opiniones se remonta al menos hasta Boole (1952). Discusiones más recientes, con copiosas referencias adicionales, incluyen a Bradley (2007), Russell, Hawthorne y Buchak (2015) y Easwaran, Fenton-Glynn, Hitchcock y Velasco (2016).

Argumentos a favor de restricciones credales

Muchas de las restricciones sobre las credibilidades presentadas en las páginas 15 a 39 tienen derecho intuitivo de considerarse exigencias racionales. Es simplemente plausible que entre más convicción tengas en que lloverá mañana, menos convicción deberías tener en que no lloverá. Pero, ¿podemos ofrecer argumentos a favor de las diversas restricciones credales? Daré aquí un panorama de tres aproximaciones históricamente importantes para argumentar a favor de restricciones sobre la credibilidad.

Argumentos de Teorema de Representación

En las páginas 46 a 48 sugerí que, si una agente tiene funciones de credibilidad y utilidad, la teoría de la decisión puede combinarlas para determinar sus preferencias racionales entre actos. Pero la teoría de la decisión también puede funcionar en la dirección opuesta. Supón que observo a una agente tomar varias decisiones a lo largo de su vida. Asumiendo que estas elecciones expresan sus preferencias entre actos, puedo construir funciones de credibilidad y utilidad para ella que racionalizarían sus preferencias si ella es una maximizadora de utilidad esperada. Podría entonces usar estas funciones de utilidad y credibilidad para predecir las elecciones que hará en el futuro.³⁵

Podemos demostrar que, mientras las preferencias de una agente sean racionales, ella puede ser representada como maxi-

35 Podemos pensar en esto como una formalización del despliegue de la gente de una “teoría de la mente”. Veo lo que haces, hago conjeturas sobre lo que quieres y lo que crees, luego dejo que esa información guíe mis interacciones contigo de ahí en adelante.

zando la utilidad esperada al combinar funciones de credibilidad y utilidad. De manera más precisa, un teorema de representación muestra que, dado un orden de preferencias sobre actos que satisface ciertos axiomas de preferencia, existe una función de utilidad y una función probabilística de credibilidad sobre las cuales esas preferencias maximizan la utilidad esperada. Puesto que hay muchas versiones distintas de la teoría de la decisión, hay muchos conjuntos de axiomas de preferencia, así que hay muchos teoremas de representación diferentes.³⁶ Pero típicamente los axiomas de preferencia suelen dividirse en dos clases: restricciones sustantivas como las exigencias de asimetría y transitividad que mencioné antes; y lo que Suppes (1974) llama “axiomas de estructura”, que especifican que el orden de preferencias es completo, tiene actos disponibles en una variedad de niveles de preferencia, etc. (usualmente se considera a los axiomas de estructura una conveniencia para hacer a los teoremas más pulcros y a las demostraciones más sencillas).

Los axiomas de representación pueden ser muy útiles. Por ejemplo, los economistas involucrados en la teoría de la elección racional a menudo modelan a los participantes del mercado como maximizando la utilidad esperada basándose en una función de utilidad y una función probabilística de credibilidad. Un teorema de representación nos asegura que, siempre y cuando una agente siga siendo racional —en el sentido de hacer elecciones que satisfagan los axiomas de preferencia—, su conducta continuará conformándose a tal modelo.

Aun así, hay un gran salto de argumentar que las agentes racionales pueden ser *modeladas* como empleando una función probabilística de credibilidad a argumentar que las agentes racionales de hecho *poseen* funciones probabilísticas de credibilidad (Hájek, 2009; Meacham y Weisberg, 2011). Podemos comenzar a ver el problema al notar que las preferencias de una agente a menudo subdeterminarán sus distribuciones de utilidad y credibilidad. Esto es, si todo lo que sabemos son las preferencias de una agente, hay (infinitamente) muchos pares distintos de funciones de utilidad y credibilidad que generarán el orden de preferencia al maximizar la

36 Los teoremas de representación se inspiraron por resultados sugerentes previos en Ramsey (1926). El primer teorema de representación riguroso del tipo de los que estamos discutiendo está en Savage (1954). (Pero véase también von Neumann y Morgenstern, 1947.) Un teorema de representación para la teoría de la decisión evidencial aparece en Jeffrey (1965), mientras Joyce (1999) demuestra uno para la teoría de la decisión causal.

utilidad esperada. Además, muchos de esos pares presentan funciones de credibilidad que no satisfacen los axiomas de probabilidad. Los teoremas de representación estándar sólo demuestran que, si las preferencias de una agente satisfacen los axiomas, *existe* un par de credibilidad/utilidad correspondiente en el cual la función de credibilidad satisface las reglas de la probabilidad. Esto difícilmente muestra que la racionalidad requiera credibilidades probabilísticas.

La situación puede mejorarse con un teorema de representación que se basa en algunas ideas a las que Lara Buchak y yo llegamos en conjunto (un esbozo de una demostración aparece en el Apéndice). Este teorema muestra que, si las preferencias de una agente satisfacen varios axiomas de preferencia, y ella maximiza utilidad esperada, entonces su función de credibilidad debe ser una *transformación escalar positiva* de una distribución de probabilidad. Con otras palabras, sus credibilidades serán no negativas, finitamente aditivas, asignarán el mismo valor a toda tautología y ese valor será mayor que el valor asignado a las contradicciones. Una función de credibilidad como ésta tendrá todas las mismas propiedades que una función probabilística, excepto que el valor máximo que asigna a las tautologías puede ser algún número positivo distinto de 1. Aun así, nada sustantivo depende de si medimos la credibilidad en una escala de 0 a 1 o, en su lugar, digamos, una escala percentil de 0 a 100.

De cualquier forma, incluso el teorema mejorado asume que las credibilidades y utilidades de la agente interactúan con preferencias a través de la maximización de la utilidad esperada. Zyn-da (2000) nota que hay muchas otras cantidades matemáticas que combinan al credibilidad y utilidad que una agente podría elegir maximizar. Así que, para argumentar a favor del probabilismo (o algo cercano a él) usando uno de estos teoremas de representación, requerimos asumir no sólo que la racionalidad exige que se satisfagan los axiomas de preferencia, sino también que exige maximizar la utilidad esperada.

Argumentos de Libro Holandés

Como con los teoremas de representación, una inspiración para los argumentos de Libro Holandés (*Dutch Book*) puede encontrarse en Ramsey (1926), en el cual él comentó

Estas son las leyes de la probabilidad, que hemos probado que son necesariamente verdaderas de cualquier conjunto consistente de grados de creencia... Si la condición mental de cualquiera violara estas leyes, su elección dependería de la forma precisa en la que se le ofrecieran las opciones, lo que sería absurdo. Podría tener una apuesta blindada hecha contra él por un apostante astuto y así estar condenado a perder en cualquier caso (p. 278).

Supóngase, por ejemplo, que estoy tanto 0.7 convencido de que iré a la oficina mañana como 0.7 convencido de que no iré. Ahora considera dos boletos de apuesta —uno que paga un dólar si voy a la oficina y otro que paga un dólar si no voy a la oficina—. Dadas mis credibilidades, mi precio de apuesta justo para cada uno de estos boletos es \$0.70. Eso significa que estoy dispuesto a pagar hasta \$0.70 por cada uno de ellos. Así que, supón que compro ambos al precio de \$0.70 cada uno. Ahora he gastado un total de \$1.40, y no importa qué suceda mañana, sólo ganaré \$1. Mi distribución de credibilidad no probabilística me ha hecho susceptible a una combinación de apuestas [en la que perderé \$0.40, sin importar qué ocurra! De Finetti (1937) demostró que, si las credibilidades de una agente violan los axiomas de la probabilidad, existe un conjunto de apuestas tal que, si la agente compra cada uno a su precio de apuesta justo, ella perderá dinero en cualquier mundo posible. Por razones desconocidas, a tal conjunto de apuestas se le conoce como un Libro Holandés (*Dutch Book*). La demostración opera yendo por cada uno de los axiomas uno a la vez, y mostrando cómo construir un Libro Holandés en contra de una agente que viola el axioma relevante. Además, podemos establecer lo que Hájek (2009) llama un Teorema Converso de Libro Holandés, mostrando que, si una agente *satisface* los axiomas de la probabilidad, ningún Libro Holandés de los tipos que figuran en la demostración de De Finetti puede construirse en contra de esa agente.

Otras demostraciones muestran cómo construir Libros Holandeses contra agentes que violan el Principio de Reflexión (van Fraassen, 1984), el Principio Principal (Howson, 1992), Regularidad (Kemeny, 1955; Shimony, 1955) y Aditividad Contable (Adams, 1962). También podemos construir lo que se conoce como una Estrategia Holandesa (*Dutch Strategy*) contra cualquier agente que viole Condicionización (Teller, 1973, reportando un resultado de David Lewis) o Condicionización de Jeffrey (Ar-

mendt, 1980; Skyrms, 1987b). Estrictamente hablando, una Estrategia Holandesa no es un conjunto particular de apuestas que garanticen dar a la agente pérdidas seguras; en su lugar, es una estrategia para hacer apuestas con la agente en las que ciertas apuestas se hacen en un momento inicial, luego se hacen apuestas futuras dependiendo de lo que la agente aprenda después de ese momento. Aun así, la idea de una Estrategia Holandesa es que sin importar qué ocurra (y sin importar qué aprenda la agente), si ella compra las apuestas a su precio de apuesta justo cuando se le ofrecen, ella enfrentará una pérdida neta pase lo que pase.

Evitar los Libros Holandeses y las Estrategias Holandesas parece ser una ventaja importante para la agente probabilística. Pero ¿podemos *sostener* que la racionalidad prohíbe ser susceptible a Libros o Estrategias Holandeses? Un problema es que los efectos negativos de violar el probabilismo enfatizados por los Libros Holandeses parecen ser peculiarmente prácticos. Podríamos haber pensado que los axiomas de Kolmogorov ofrecían restricciones de la racionalidad *teórica* (más que *práctica*) sobre las credibilidades de las agentes. Pero aquí estamos argumentando a favor de esos axiomas al indicar las consecuencias financieras de violarlos. Además, no es claro qué tan en serio debemos tomarnos esas consecuencias potenciales. ¿Realmente enfrentarán alguna vez las agentes no probabilísticas el conjunto preciso de apuestas que las expondrían a un Libro Holandés? ¿Y qué tal si la agente no probabilística ha leído sobre Libros Holandeses y decide que, en lugar de cambiar sus credibilidades, sólo será más cuidadosa en su conducta de apuestas? En el ejemplo anterior sobre ir a la oficina, yo podría pagar \$0.70 por la apuesta que me da ganancias si voy a la oficina, pero luego negarme a comprar la segunda apuesta debido a que veo venir un Libro Holandés. En ese caso aun tendré credibilidades no probabilísticas, pero me las arreglaré con estrategias prácticas para evitar los prospectos de pérdidas seguras.

Siguiendo una pista de la segunda oración de la cita anterior de Ramsey, varios autores han intentado “despragmatizar” los argumentos del Libro Holandés. Skyrms escribe que “para Ramsey, el apostador astuto es un dispositivo *dramático* y la posibilidad de un Libro Holandés es un *síntoma* impresionante de una incoherencia más profunda” (Skyrms, 1987a, p. 277, cursivas mías). Para estos autores,³⁷ la susceptibilidad a los Libros Holandeses

37 Véase también Armendt (1992), Christensen (2004) y Howson y Urbach (2006).

meramente saca a la luz una inconsistencia subyacente en las credibilidades de la agente –la inconsistencia de evaluar la misma cosa de diferentes maneras, dependiendo de cómo sea presentada–.

Volvamos a mis apuestas sobre si iré a la oficina mañana. Dada mi convicción de 0.7 en que iré, mi precio de apuesta justo para una apuesta que paga \$1 si voy y nada de otra manera es \$0.70. Así que valúo esa apuesta en \$0.70; si se me ofrece la oportunidad de comprar esa apuesta a cualquier precio inferior –digamos, \$0.50– consideraré eso un trato favorable. Por otra parte, mi convicción de 0.7 en que no iré me da un precio de apuesta justo de \$0.70 por una apuesta que paga \$1 si *no voy* y nada si voy. Así que consideraría poco favorable vender esa apuesta a cualquier precio menor que \$0.70 –por ejemplo, \$0.50–. Aun así, comprar la primera apuesta a \$0.50 y vender la segunda apuesta en \$0.50 *son exactamente la misma transacción*; cada una me daría \$0.50 netos si voy a la oficina y me haría perder \$0.50 si no voy. Así que, ¿veo la transacción favorablemente o no? Una de mis credibilidades sugiere que la veo favorablemente, mientras la otra exige que no. Cómo esas credibilidades evalúan esas apuestas revela el conflicto entre ellas.³⁸

Pero incluso los argumentos del Libro Holandés despragmatizados hacen supuestos potencialmente controversiales. Primero, asumimos que los precios de apuesta justos de una agente racional son iguales a sus ganancias esperadas –una suposición que podría no cumplirse en agentes con aversión al riesgo–. Y segundo, para construir un Libro Holandés contra violaciones de Aditividad Finita, necesitamos asumir un “principio de paquete” –que el precio de apuesta justo de una agente racional para una combinación de dos apuestas es igual a la suma de sus precios de apuesta para cada una considerada individualmente–. Cada una de estas suposiciones se seguiría fácilmente si asumiéramos que las agentes racionales siempre eligen maximizar la utilidad esperada. Pero si pudiéramos asumir *eso*, ya tendríamos un argumento de teorema de representación para algo muy cercano al probabilismo (en páginas 51 a 53).³⁹

38 Nota que no tendría este problema si satisficiera el cálculo de probabilidades al, digamos, asignar credibilidad 0.7 a que iré y 0.3 a que no iré. En ese caso buscaré favorablemente comprar la primera apuesta a \$0.50 y también buscaré favorablemente vender la segunda en \$0.50, así que mis evaluaciones estarán en armonía.

39 De hecho, la demostración del teorema de representación en el Apéndice refleja a detalle la estructura de los teoremas tradicionales del Libro Holandés a favor del probabilismo.

Así que es poco claro para qué se requeriría la desviación a través de apostadores astutos.

Argumentos de Precisión

James M. Joyce (1998) se propone proporcionar una “reivindicación no pragmática del probabilismo” que evite explícitamente invocar consecuencias prácticas en su defensa de los axiomas de la probabilidad como restricciones racionales sobre la credibilidad. Su obra parte de resultados matemáticos que se encuentran en De Finetti (1974) y Rosenkrantz (1981), pero los usa para construir una nueva clase de argumento.

La idea clave de Joyce es que, desde el punto de vista de la racionalidad teórica pura, las agentes deberían proponerse hacer que sus credibilidades sean tan precisas como sea posible. ¿Cómo podríamos medir la precisión de una función de credibilidad? Históricamente, una opción había sido centrarse en la calibración. La función c está perfectamente calibrada si, para todo $0 \leq x \leq 1$, cuando vemos todas las proposiciones en $\mathcal{L}$ a las que c asigna credibilidad x , la fracción de esas proposiciones que son verdaderas es exactamente x . Si estoy perfectamente calibrado, exactamente la mitad de las proposiciones a las que asigno credibilidad $1/2$ son verdaderas, exactamente un tercio de las proposiciones a las que asigno credibilidad $1/3$ son verdaderas, etc.

Van Fraassen (1983) y Shimony (1988) argumentan a favor del probabilismo al mostrar que, para que una distribución de credibilidad sea incrustable en sistemas más y más grandes que se acercan a la calibración perfecta, esa distribución de credibilidad debe satisfacer los axiomas de la probabilidad. Esto podría fungir como un buen argumento a favor del probabilismo, pero la calibración tiene algunos rasgos intuitivamente indeseables como medida de precisión. Por ejemplo, considera a dos agentes que asignan credibilidades a cuatro proposiciones como en la Tabla 1.

Tabla 1. Dos asignaciones de credibilidad

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
AGENTE 1	0.5	0.5	0.5	0.5
AGENTE 2	1	1	0.01	0
VALORES DE VERDAD	V	V	F	F

Espero aceptes que, intuitivamente, las credibilidades de Agente 2 son mucho más precisas (cercanas a la verdad) que las de Agente 1. Pero Agente 1 está perfectamente calibrada –exactamente la mitad de las proposiciones a las que asigna credibilidad $\frac{1}{2}$ son verdaderas– mientras que Agente 2 no lo está.

Nuestras intuiciones sobre precisión funcionan viendo cada asignación de credibilidad una a la vez, evaluando qué tan precisa es esa credibilidad dado el valor de verdad de la proposición y, luego, agregando esas evaluaciones locales de precisión a través de todas las proposiciones. Pero la calibración funciona con rasgos *globales* de una distribución de probabilidad, lo cual (como hemos visto) puede llevar a efectos distorsionadores.

Así que Joyce usa una aproximación gradual a la precisión en su lugar. En esta aproximación, seleccionamos una regla de puntuación para medir qué tan lejos está cada asignación de credibilidad individual a una proposición de la verdad acerca de esa proposición. Intuitivamente, cuando la proposición *P* es verdadera, las credibilidades en *P* más altas son más precisas; cuando *P* es falsa, las credibilidades más bajas son mejores. Podemos formalizar esto al tener una función *I* que asigna 1 a *P* si es verdadera y 0 si es falsa, luego midiendo cuán lejos está $c(P)$ de $I(P)$. Históricamente, ha sido popular medir esta distancia como

$$(I(P) - c(P))^2. \tag{24}$$

Nota que este resultado de medición *se incrementa* entre *más lejos* te encuentres de la verdad; así que es una medida de *imprecisión* credal. Una agente racional que se propone ser tan precisa como sea posible debería intentar minimizar esta cantidad para cada proposición. Globalmente, debería intentar minimizar la suma de esta cantidad a través de todas las proposiciones que con-

sidera (comúnmente se conoce a esta suma como puntuación de Brier, en honor de la discusión del meteorólogo George Brier en “Verification of forecasts expressed in terms of probability”, 1950).

Joyce muestra que si usamos la puntuación de Brier para medir precisión, entonces cualquier distribución de credibilidad no probabilística será dominada en precisión por otra distribución, que será probabilística, sobre el mismo conjunto de proposiciones. Esto es, si consideras a una agente cuyas credibilidades sobre algún lenguaje violan los axiomas de la probabilidad, habrá otra distribución de credibilidad sobre el mismo lenguaje que sea probabilística la cual tendrá una puntuación de Brier más precisa que la de ella *en todo mundo posible*. Cuando la agente no probabilística considere esa distribución alternativa, sabrá que es más precisa que la suya, incluso sin saber nada sobre cuál mundo posible es el real. Joyce sostuvo que sería irracional que una agente mantuviera su distribución no probabilística, a pesar de disponer de esta información de que otra distribución era con certeza más precisa. Y puesto que cualquier agente cuyas credibilidades violen los axiomas de la probabilidad enfrentará la misma situación, esto constituye un argumento a favor del probabilismo.⁴⁰

Se han ofrecido argumentos de precisión emparentados a favor de varias otras normas bayesianas: Condicionización (Greaves y Wallace, 2006; Briggs y Pettigrew, 2020), el Principio Principal (Pettigrew, 2013), el Principio de Indiferencia (Pettigrew, 2014), Reflexión (Easwaran, 2013) y Conglomerabilidad (Easwaran, 2013).

Hay dos inquietudes principales en la literatura sobre estos argumentos de precisión. Primero, hay una inquietud general sobre evaluar la racionalidad de las credibilidades al medir su distancia de la verdad. La aproximación gradual revela una clase de consecuencialismo epistémico, en el cual las actitudes apuntan a algún resultado (en este caso, la verdad), y son evaluadas según qué tan bien se aproximan a ese objetivo. Tal como las aproximaciones teleológicas a la normatividad han despertado sospecha en la ética y otras áreas de la filosofía, el programa de precisión gradual ha sido criticado por autores como Greaves (2013), Berker (2013) y Carr (2017).

Segundo, entre quienes aceptan el programa de precisión gradual, hay una inquietud sobre cómo seleccionar una regla de

40 De manera importante, no puede ofrecerse la misma clase de argumento en contra del probabilismo. Una función de credibilidad que satisface los axiomas de la probabilidad no será dominada en precisión de la manera que Joyce describe por otra función (probabilística o no).

puntuación para medir precisión. Maher (2002) sugiere que en vez de usar la puntuación de Brier, podríamos estimar la distancia entre una credibilidad individual $c(P)$ y un valor de verdad $I(P)$ al calcular

$$|I(P) - c(P)|. \quad (25)$$

Históricamente, se prefirió la puntuación de Brier por encima de esta puntuación de valor absoluto debido a que la primera es una regla de puntuación “apropiada” mientras que la segunda no. Para entender la diferencia, supón que se ha lanzado un dado de seis lados y que tenemos a dos personajes que no conocen el resultado. Nuestro primer personaje, Suertuda, asigna credibilidad $1/6$ a cada uno de los posibles resultados. Nuestro segundo personaje, Pesimista, asigna credibilidad 0 a cada resultado. La función de credibilidad de Suertuda satisface los axiomas de la probabilidad, mientras que la de Pesimista no.

Ahora supón que cada uno de nuestros personajes calcula un valor de imprecisión esperada para sí misma y para la otra persona. Para dar un ejemplo de cómo funciona esto, supón que Suertuda calcula un valor de imprecisión esperada para su propia distribución usando la puntuación de Brier. Para hacerlo, Suertuda considera cada uno de los seis mundos posibles disponibles (esto es, cada uno de los seis posibles resultados del lanzamiento del dado), evalúa cuál sería su puntuación de Brier en ese mundo posible, multiplica por su credibilidad de que ese mundo posible sea el real, luego suma a través de todas las posibilidades. Si, por ejemplo, el lanzamiento del dado resulta en 3, su puntuación de Brier será

$$\begin{aligned} & (I(1) - c(1))^2 + (I(2) - c(2))^2 + (I(3) - c(3))^2 + (I(4) - c(4))^2 + (I(5) - c(5))^2 + (I(6) - c(6))^2 = (0 - 1/6)^2 \\ & + (0 - 1/6)^2 + (1 - 1/6)^2 + (0 - 1/6)^2 + (0 - 1/6)^2 + (0 - 1/6)^2 \end{aligned} \quad (26)$$

$$= 1/36 + 1/36 + 1/36 + 25/36 + 1/36 + 1/36 = 30/36 = 5/6.$$

Un poco de reflexión mostrará que ésta es la puntuación de Brier de Suertuda en cada uno de los seis mundos posibles. Así que su puntuación de Brier a través de todos esos mundos es también $5/6$. Mientras tanto, dejaré en manos del lector calcular que la pun-

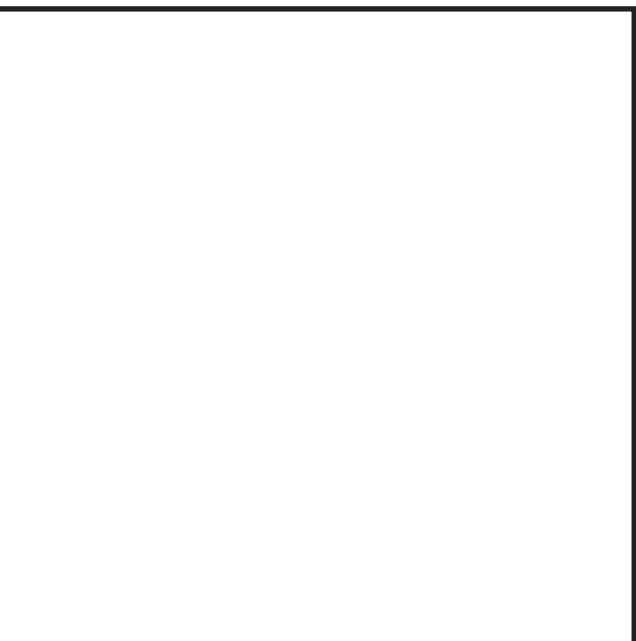
tuación esperada de Brier de Pesimista es 1. Puesto que sus puntuaciones más altas significan mayor imprecisión –y menor precisión–, Suertuda espera que sus credibilidades sean más precisas que las de Pesimista cuando se usa la puntuación de Brier para calcular precisión.

Ocorre exactamente lo opuesto si usamos la medida de valor absoluto. De nuevo, dejaré al lector calcular que la puntuación de valor absoluto esperado de Suertuda es $5/3$, mientras que el de Pesimista es de nuevo 1. Así que, a la luz de la puntuación de valor absoluto, se espera que la no probabilística Pesimista sea más precisa que la probabilística Suertuda.

Las reglas de puntuación apropiadas son reglas en las que una agente probabilística nunca esperará que algún otro agente sea más preciso que ella. La puntuación de Brier es una de las muchas reglas de puntuación apropiadas, mientras la puntuación de valor absoluto es inapropiada. En general, parece irracional que una agente se aferre a una distribución de credibilidades cuando espera que las credibilidades de otra agente sean más precisas que las suyas (Lewis, 1971). Así que una teórica que ya ha aceptado que las distribuciones probabilísticas son racionales tiene buenas razones para trabajar con reglas de puntuación apropiadas en vez de con reglas inapropiadas. Los argumentos basados en precisión a favor de Condicionización, el Principio Principal, el Principio de Indiferencia, etc., antes mencionados se confinan todos a trabajar con reglas de puntuación apropiadas.

Predd *et al.* (2009) muestran que el argumento de dominancia de la precisión de Joyce a favor del probabilismo podría formularse usando cualquier regla de puntuación apropiada. Pero en el contexto de un argumento a favor del probabilismo, dar preferencia a las reglas de puntuación apropiadas sobre las inapropiadas parece una petición de principio. Se define a las reglas de puntuación apropiadas como aquellas en las que las distribuciones probabilísticas son calificadas como esperadamente más precisas que las alternativas. A menos que tengas una razón antecedente para pensar que las distribuciones probabilísticas deberían resultar más prometedoras que las alternativas, ésta no es una razón para preferir una puntuación apropiada.⁴¹

41 Aunque podría haber otras razones. Véase, *e.g.*, Joyce (2009) y Pettigrew (2016).



Argumentos en contra de restricciones credales

Tras revisar algunos argumentos a favor de varias restricciones racionales sobre las credibilidades, ¿cuáles son los argumentos en contra de estas restricciones? Hay muchos, por supuesto, y se multiplican con el tiempo. Aquí me centraré en unos cuantos que han generado discusión profunda e interesantes respuestas positivas.

El Problema de la Omnisciencia Lógica

Savage (1967) famosamente consideró el apuro de “una persona a la que se le solicita arriesgar dinero sobre un dígito remoto de π ”. Su preocupación era que, de acuerdo con el axioma de Normalidad, se requiere que una agente asigne certeza a cualquier tautología en su lenguaje $\mathcal{L}$. Presumiblemente, el hecho de que un dígito específico de π tenga un valor particular es una tautología.⁴² Así que, de acuerdo con el probabilismo, una agente racional debería estar segura de todos los dígitos de π . Pero esto parece demasiado para que la racionalidad lo exija de cualquier agente real.

La discusión de Savage inició una literatura sobre lo que se conoce como el Problema de la Omnisciencia Lógica. De hecho, yo creo que hay múltiples problemas relacionados aquí, que podríamos etiquetar como sigue.⁴³

42 Si tus puntos de vista sobre el logicismo en la filosofía de las matemáticas tienen como consecuencia lógica que los hechos sobre los dígitos de π no son tautologías, siempre podemos substituirlos en un condicional cuyo antecedente sean varios axiomas matemáticos y cuyo consecuente reporte un dígito de π . O podemos usar en su lugar algunas verdades lógicas muy complejas.

43 La etiqueta “Aprendizaje lógico” es común en la literatura; yo inventé las otras dos etiquetas para nuestra presente discusión.

COMPLETUD CREDAL

El probabilismo requiere que una agente le asigne una credibilidad a cada proposición en su lenguaje.

DISCERNIMIENTO LÓGICO

El probabilismo prohíbe que una agente asigne una credibilidad distinta de 1 a cualquier tautología.

APRENDIZAJE LÓGICO

Una agente probabilística nunca pasará de una credibilidad baja en una tautología a una credibilidad más alta.

El problema de la Completud Credal es que los axiomas de la probabilidad requieren que una agente asigne una credibilidad a toda proposición en su lenguaje. Por ejemplo, No-Negatividad dice que cualquier $X \in \mathcal{L}$ recibe algún valor no negativo de credibilidad. Incluso en un lenguaje con una cantidad finita de proposiciones atómicas, la clausura bajo conectivas veritativo funcionales generará un lenguaje de tamaño infinito. Pero parece que no sólo es imposible que una agente finita asigne tantas credibilidades, sino también que es poco recomendable bajo el principio de Abstención de Acumulación Innecesaria (*Clutter Avoidance*), de Harman (1986).

ABSTENCIÓN DE ACUMULACIÓN INNECESARIA

Uno no debería abarrotar la propia mente con trivialidades.

Pero podemos alterar ligeramente nuestro formalismo de modo que deje de exigir Completud Credal y evada inquietudes sobre Abstención de Acumulación Innecesaria. La idea es exigir, no que una distribución de credibilidades de la agente de hecho satisfaga los axiomas de la probabilidad, sino sólo que sea *extendible* a una distribución que lo hace. Con otras palabras, permitimos que una agente adopte una distribución de credibilidades parcial que asigna valores numéricos a sólo algunas de las proposiciones en $\mathcal{L}$, pero exigimos que haya alguna manera posible de asignarle valores al resto de $\mathcal{L}$, de modo que la distribución resultante completa satisfaga los axiomas. Esta aproximación recupera resultados intuitivos tales como la restricción de que, si una agente asigna credibilidades tanto a P como a $\neg P$, esas credibilidades deben sumar 1. Pero no culpará a una agente si no adopta actitudes hacia P , $\neg P$, o ambas.

Pasar a distribuciones parciales evita el problema de la Completud Credal, pero deja intacto el problema del Discernimiento Lógico. Parece perfectamente racional que yo asigne credibilidad de $1/10$ a que el billonésimo dígito de π es 2. Pero cualquier distribución de credibilidad —parcial o completa— que contenga esa asignación no es extendible a una distribución probabilística. Es una tautología que el billonésimo dígito de π es 2, o es una tautología que el billonésimo dígito de π no es 2, así que el probabilismo exige que yo asigne a esa proposición una credibilidad de 1 o exige que le asigne una credibilidad de 0. Cualquiera que sea la verdadera exigencia, parece demasiado demandante, puesto que no tengo una buena manera de descifrar cuál es la exigencia.

Antes de considerar respuestas a este problema de Discernimiento Lógico, consideremos rápidamente Aprendizaje Lógico. La siguiente secuencia credal parece bastante razonable: le asigno credibilidad de $1/10$ a que el billonésimo dígito de π es 2, Talbott (1991) me dice que de hecho es 2, así que mi credibilidad de que es 2 dramáticamente se incrementa (quizá hasta llegar a 1). En este caso parece que he aprendido una verdad lógica y mi incremento credal es una respuesta racional a ese episodio de aprendizaje. Pero un sistema bayesiano tradicional no aprobará esta respuesta, o no será capaz de modelarla de manera útil, puesto que un sistema probabilístico sólo permite distribuciones de credibilidad (en cualquier momento) que asignan a esa proposición un valor de 1.

Si resolviéramos el problema del Discernimiento Lógico al construir una teoría bayesiana que permitiera credibilidades racionales en tautologías distintas de 1, presumiblemente esa teoría también permitiría incrementos y disminuciones en tales credibilidades. Así que hay esperanza de que una solución a Discernimiento Lógico haga posible una solución a Aprendizaje Lógico.

¿Cómo, entonces, podríamos modelar a una agente bayesiana sin discernimiento lógico perfecto? En respuesta a Savage, Hacking (1967) sugiere que identifiquemos una proposición como “personalmente posible” para una agente si la agente no sabe que es falsa. Luego ajustamos Normalidad para exigir certeza sólo en proposiciones cuyas negaciones son personalmente imposibles, y Aditividad Finita para que se aplique sólo cuando $P \& Q$ es personalmente imposible. Esto permite que una agente sea ignorante sobre una cantidad arbitrariamente amplia de verdades lógicas y, por tanto, esté menos que segura de esas verdades.

Pero esta aproximación crea tres problemas. El primero es formal. Hacking trabaja con distribuciones de credibilidades sobre oraciones y es libre de tratar las oraciones que él quiera como personalmente posibles o imposibles. Pero si pensamos en esas oraciones como representando proposiciones subyacentes, y a esas proposiciones como representando conjuntos de posibilidades subyacentes, parece natural preguntarse qué posibilidades contempla una agente cuando considera como personalmente posible aquello que es lógicamente imposible. Para atender esta clase de hueco, Hintikka (1975) construye una semántica que admite mundos lógicamente imposibles, los cuales pueden entrar en el contenido de proposiciones de la misma manera que los mundos posibles clásicos.

Un segundo problema intuitivo es que la aproximación de Hacking permite cantidades arbitrariamente amplias de no omnisciencia lógica —nada en el formalismo de Hacking condena a una agente que asigna menos que certeza a $P \vee \neg P$, mientras que la agente no sepa que la proposición es verdadera—. Bjerring y Skipper (2022) reclaman que el formalismo de Hacking es tan permisivo que, al sacrificar omnisciencia lógica, no logra capturar ningún requerimiento racional de competencia lógica básica. Expresan inquietudes similares sobre un marco conceptual a partir de Garber (1983), y varios formalismos desarrollados usando la semántica de Hintikka.

Finalmente, es importante ver qué pierde un sistema bayesiano cuando es redefinido en términos de posibilidad personal en lugar de lógica. Si una agente no logra saber que $P \& \neg P$ es imposible, entonces a la luz de Hacking no necesita aplicar Aditividad Finita a P y $\neg P$. Como resultado, tal agente podría asignar credibilidades a P y $\neg P$ que sumen más de 1. Ella puede incrementar su credibilidad en P sin disminuir su credibilidad en $\neg P$. En nuestra teoría de la confirmación basada en relevancia, ella puede no ver a P como socavando $\neg P$. Y cuando ella selecciona acciones al maximizar la utilidad epistémica esperada, ella puede violar los axiomas de preferencia de varias maneras. Con otras palabras, las características y aplicaciones mismas que hacen del bayesianismo una imagen plausible de la racionalidad comienzan a disolverse una vez que se relajan las restricciones de discernimiento lógico.

Así que ¿quizá deberíamos ir en la otra dirección? Varios teóricos han comenzado a preguntarse si los requerimientos de omnisciencia lógica no son un efecto colateral de nuestros formalismos epistémicos, sino en su lugar un indicio por parte de esos

formalismos sobre el dominio normativo subyacente. Smithies (2015) sostiene que la certeza en verdades lógicas es de hecho un requerimiento de la racionalidad; Titelbaum (2015) y Littlejohn (2018) abogan por posiciones similares.

El Problema de la Vieja Evidencia

Clark Glymour inició el debate sobre la Vieja Evidencia con un famoso ejemplo.

Los científicos habitualmente argumentan a favor de sus teorías a partir de evidencia que se conoce desde mucho antes de que sus teorías fuesen introducidas... El argumento que Einstein dio en 1915 a favor de sus ecuaciones de campo gravitacional fue que explicaban el anómalo avance del perihelio de Mercurio, que había sido determinado más de medio siglo antes. A otros físicos el argumento les pareció enormemente convincente, y es una conjetura justa que sin él los británicos no habrían organizado la famosa expedición del eclipse de 1919. La vieja evidencia puede de hecho confirmar una nueva teoría, pero de acuerdo con la cinemática bayesiana, no puede hacerlo (Glymour, 1980, pp. 306-7).

Ya hemos visto (páginas 24 a 31) que un bayesiano tradicional modela la adquisición de evidencia como la obtención de certezas, las cuales luego son retenidas. Al mismo tiempo (páginas 41 a 46), se entiende la confirmación como relevancia positiva. Combinando estas dos aproximaciones, tenemos un problema: una vez que se ha aprendido una proposición evidencial, recibe credibilidad 1. Cuando $c(E) = 1$, $c(H|E) = c(H)$ para cualquier $H \in \mathcal{L}$. Así que, una vez que una agente aprende algo, esa pieza de información es por siempre confirmacionalmente inerte de ahí en adelante.

Dados estos hechos básicos sobre el bayesianismo, podemos identificar dos desafíos en la historia de Glymour sobre Einstein. Christensen (1999) los llama los problemas “sincrónico” y “diacrónico” de la vieja evidencia.⁴⁴ El problema diacrónico es sobre el cambio en credibilidad. En el curso de 1915, Einstein incrementó

44 Uso la terminología de Christensen debido a que encuentro que es la más útil. Pero antes, desambiguaciones similares del problema de la vieja evidencia pueden encontrarse en Garber (1983), Eells (1985) y Zynda (1995).

su convicción en la Teoría General de la Relatividad (TGR), y creemos que esto tuvo algo que ver con el perihelio de Mercurio. Pero no puede ser que Einstein incrementara su convicción debido a que aprendió sobre su avance anómalo —él ya sabía sobre eso mucho antes de 1915—. Así que, ¿qué cambió su opinión y cómo podemos reflejarlo en un sistema bayesiano?

El problema sincrónico de la vieja evidencia aparece después de 1915, cuando el perihelio de Mercurio ya hizo efecto sobre las actitudes de Einstein hacia TGR. Presumiblemente, incluso después de 1915, Einstein habría citado el avance del perihelio de Mercurio como una pieza crucial de evidencia que apoya TGR. Pero relativo a la función de credibilidad de Einstein en ese momento —la cual asigna 1 a hechos del perihelio— esos hechos no son positivamente relevantes para TGR. Así que, ¿cómo puede un bayesiano sobre la confirmación interpretar ese apoyo evidencial?

Las propuestas para resolver el problema sincrónico usualmente funcionan relativizando la confirmación con alguna función de probabilidad distinta de la distribución actual de credibilidades de la agente. Puesto que la agente actualmente asigna $c(E) = 1$, E no puede confirmar nada relativo a esa distribución actual. Así que buscamos alguna otra distribución relevante que no asigne 1 a E . Por ejemplo, podríamos adoptar una aproximación de “rastreo retrospectivo histórico”, en la cual miramos a algún momento previo en el que la agente aun no estaba segura de E , y nos preguntamos si E era positivamente relevante para H en su distribución de credibilidades en ese momento. Pero esta aproximación es limitada por varias razones. Por ejemplo, Einstein probablemente supo sobre el perihelio de Mercurio mucho antes de siquiera considerar TGR. Así que, si hacemos un rastreo retrospectivo a un momento muy anterior a 1915 en el que Einstein aun no estaba seguro de E , no seremos capaces de encontrar ninguna credibilidad condicional o incondicional que él asignara a la H relevante en ese momento. Por ende, no seremos capaces de decir que E confirma H para Einstein, debido a que en algún momento en el pasado él asignó $c(H|E) > c(H)$.

A la luz de éstas y otras dificultades, Howson y Urbach (2006) proponen una aproximación de “rastreo retrospectivo contrafáctico”. En lugar de ver a un momento del pasado en el que la agente no conocía E , vemos hacia un mundo posible cercano en que la agente sabe todo lo que ella sabe excepto E . Bueno, no exactamente *todo* —probablemente también queremos un mundo

en el que ella no conoce los equivalentes lógicos de E , las consecuencias lógicas inmediatas de E , etcétera—. Pero Howson y Urbach (p. 300) tienen una propuesta técnica para identificar las proposiciones que deberían sustraerse. Dejando de lado los detalles técnicos, a Earman (1992, p. 123) le preocupa que esta aproximación contrafáctica sufrirá de defectos similares a los de otros análisis contrafácticos; desplazarse a un mundo no real puede tener efectos colaterales que echen a perder el análisis. Por ejemplo, el registro histórico sugiere que Einstein fue motivado para formular TGR en parte para explicar el avance anómalo de Mercurio. Así que el mundo posible más cercano en el que Einstein no conoce E , pero aun así le asigna credibilidades a H puede estar de hecho muy lejos —y ser muy diferente— del nuestro.

Quizá la mejor aproximación sea decir que cuando una agente explica la evidencia que apoya alguna hipótesis, el apoyo que describe puede ser relativo no a sus propias credibilidades personales sino a las de alguna otra distribución probabilística. Esa distribución puede ser una que su audiencia, o una comunidad científica particular, asume pertinente. O si somos bayesianos objetivos (página 31), puede ser la distribución objetiva que determina cómo deberían establecer sus credibilidades todas las agentes racionales. Maher (1996), por ejemplo, desarrolla una propuesta del último tipo. Pero quedan aún por resolver muchos detalles. Por ejemplo, ¿cómo una comunidad científica o una distribución racional objetiva asignan una probabilidad inicial a la proposición de que TGR expresa las leyes físicas para nuestro universo?⁴⁵

En cuanto al problema diacrónico de la vieja evidencia, la respuesta típica es identificar algo *distinto* al aprendizaje del avance del perihelio de Mercurio que le diera a Einstein nueva convicción en TGR durante el trascurso de 1915. Por poner un caso, Einstein podría haber descubierto en algún momento de 1915 no que el perihelio de Mercurio avanza anómalamente, sino que TGR predice tal avance anómalo. Puesto que es un hecho lógico que TGR (junto con otra información empírica de la que Einstein ya estaba consciente) tiene como consecuencia lógica los detalles del avance, éste sería un caso de aprendizaje lógico. Así que una implementa-

45 Incluso si quisiéramos usar un Principio de Indiferencia (página 32) aquí, necesitaríamos una partición sobre la cual dividir equitativamente nuestra credibilidad, y es difícil determinar qué conjuntos alternativos de leyes físicas deberían entrar en tal partición.

ción bayesiana de esta explicación dependerá de las cuestiones sobre la omnisciencia lógicas discutidas en las páginas 63 a 66.

Otra posibilidad es que la elevada convicción de Einstein en TGR al final de 1915 fuese nueva debido a que él no tenía *ninguna* actitud hacia TGR al inicio de 1915. Quizá Einstein aun no concebía TGR a inicios de 1915, así que el lenguaje sobre el que asignaba credibilidades en ese momento no contenía una proposición que expresara la verdad de TGR. Esta aproximación ciertamente explicaría por qué Einstein tenía una nueva elevada credibilidad al final de ese año que él no tenía al inicio. Pero probablemente no se generaliza a todos los casos de confirmación mediante vieja evidencia (y puede que no sea históricamente precisa en el caso de Einstein). Además, casos en los que agentes añaden nuevas proposiciones a su lenguaje cognitivo plantean otro desafío al bayesianismo. Todas las normas de actualización que hemos considerado (Condicionización, Condicionización de Jeffrey) operan sobre un lenguaje que se mantiene fijo a través del tiempo. El denominado “problema de las nuevas teorías” nos reta a construir un formalismo que permita que el lenguaje de una agente cambie con el tiempo, y que ponga restricciones razonables sobre cómo las credibilidades de la agente deberían evolucionar a través de tales cambios.

Finalmente, podríamos enfocarnos en el hecho de que ambas versiones del Problema de la Vieja Evidencia parecen surgir debido a que Condicionización trata la adquisición de evidencia como obtener certezas. Si la evidencia novedosamente adquirida no fuese a (y permaneciera en) una credibilidad de 1, entonces no tendríamos el problema de que la vieja evidencia siempre tiene credibilidad 1 y, por ende, no puede ser positivamente relevante para nada. Supóngase que adoptamos el Principio de Regularidad (que prohíbe certeza sobre proposiciones empíricas), y decretamos que Condicionización de Jeffrey es el esquema de actualización racional. Entonces la adquisición de evidencia incrementará la credibilidad en proposiciones particulares, pero nunca la llevará a 1, y nunca surgirá el Problema de la Vieja Evidencia.

Christensen (1999) elabora esta aproximación y encuentra mucho que la recomienda, pero eventualmente encuentra una nueva dificultad. El Problema de la Vieja Evidencia es que adquirir una pieza de evidencia no debería privarla de su habilidad para confirmar hipótesis. Generalizando esta idea, deberíamos concordar en que llegar a estar más convencido en una pieza de evidencia no debería afectar el grado en el que confirma una hipótesis. Así

que Christensen busca una medida de confirmación (páginas 44 a 46) en la que las Condicionalizaciones de Jeffrey que cambian $c(E)$ no afecten el nivel de confirmación de E sobre H . No logra encontrar una medida que satisfaga esta restricción, reúna otras condiciones formales plausibles y funcione intuitivamente en ejemplos.

Pérdida de Memoria y Sensibilidad al Contexto

La retención y adquisición de certeza también plantean otros problemas para el marco conceptual de actualización basada en Condicionización. Por ejemplo, muchos de nosotros tenemos la experiencia de obtener una pieza de evidencia un día y luego olvidarla poco tiempo después. Pero si somos condicionalizadores constantes, una proposición que alcanza credibilidad 1, en algún momento, nunca puede caer en picada a una credibilidad menor después. Así que Condicionización juzga irracional la pérdida de memoria.⁴⁶

Aunque el reconocimiento de este problema se remonta al menos hasta Levi (1987), Talbott (1991) lo presenta de manera particularmente contundente. Él considera la respuesta de que las reglas bayesianas se proponen modelar agentes idealmente racionales —no agentes cotidianas— “y una agente idealmente racional no estaría sujeta a la falibilidad de la memoria humana” (p. 141). Si sirve de algo, no veo por qué la memoria paquidérmica haría a una agente más racional que otra (aunque véase Carr, 2015), pero el asunto entero puede ser esquivado por medio de un ingenioso ejemplo que se debe a Arntzenius (2003). Aunque no elaboraré los detalles aquí, la conclusión del ejemplo de Arntzenius es que Condicionización no sólo condena a agentes que de hecho olvidan evidencia, sino también a agentes que sospechan que podrían haber olvidado evidencia (incluso si en realidad no lo han hecho). Seguramente no podemos exigir de agentes idealmente racionales

46 O, por lo menos, la versión de Condicionización que hemos estado discutiendo considera irracional la pérdida de memoria, debido a que gobierna la actualización de una agente a través de cualquier intervalo arbitrario de momentos t_i a t_j . Podría adoptarse una versión más limitada de Condicionización (compárese con Titelbaum, 2013a, Cap. 6) que se aplique sólo a través de intervalos durante los cuales la información de la agente se incrementa estrictamente. En ese caso, el problema sería que más que juzgar irracional la pérdida de memoria, la regla de actualización limitada no logra darnos ninguna orientación en casos de pérdida de memoria.

certeza en la proposición empírica de que ¡nunca han olvidado nada en sus vidas!

¿Podemos alterar Condicionalización para permitir pérdida de certeza? Una aproximación popular es sacar ventaja de un rasgo que ya exhibe Condicionalización tradicional. Supón que tenemos una agente que condicionaliza a lo largo de su vida entera. A medida que obtiene evidencia, ella irá acumulando certezas; el conjunto total de certezas que ella posee en cualquier momento representará su evidencia total en ese momento. Refirámonos a la proposición que expresa la conjunción de toda la evidencia/certezas de la agente en el momento t_i como E_i . Si la agente es una condicionalizadora devota, habrá al menos una distribución de probabilidad regular⁴⁷ p_h tal que para cualquier momento t_i en el que la agente asigna credibilidades, y cualquier proposición X en su lenguaje $\mathcal{L}$, $c_i(X) = p_h(X|E_i)$. Con otras palabras, existe una única función p_h vinculada a cada momento en la vida de la agente, tal que su distribución de credibilidad en cualquier momento puede recuperarse al condicionalizar p_h sobre su evidencia total en ese momento.

Me referiré a esta distribución p_h como el *inicial hipotético* de la agente; a veces también se le llama “*ur-inicial*” (*ur-prior*) o una “distribución de credibilidad inicial”. Este último alias proviene de pensar en p_h como representando las credibilidades de la agente en algún momento anterior en su vida cuando carecía de cualesquiera certezas empíricas. Debido a que la condicionalización es acumulativa y conmutativa, si una agente *tuvo* un momento inicial tal en su vida —antes de su primera actualización por Condicionalización—, las credibilidades que asignaba en ese momento *se vincularían* con sus opiniones posteriores de la manera en que p_h se vincula con c_i . Pero es difícil imaginar que cualquier agente real haya tenido un momento en el que careciera por completo de información empírica.

Así que prefiero pensar en el inicial hipotético de una agente como una herramienta conveniente para separar dos influencias sobre sus credibilidades. Por una parte, está su evidencia; por la otra, están sus estándares epistémicos, los cuales encapsulan sus principios y tendencias para interpretar la evidencia. La evidencia total de la agente cambia a través del tiempo y se representa en el momento t_i por E_i . Pero, a medida que su

47 Al decir que la distribución es “regular,” queremos decir que asigna credibilidad 1 sólo a verdades lógicas.

evidencia cambia, ella puede retener un conjunto constante de estándares para interpretar la evidencia, representado por su inicial hipotético p_h . Aplicar estos estándares a la evidencia total de la agente en t_i –al condicionalizar p_h sobre E_i – produce su distribución de credibilidad en c_i .⁴⁸

Esta imagen atractiva en general es consecuencia lógica de Condicionalización: si una agente condicionaliza en cada actualización, *entonces* sus credibilidades a lo largo de su vida serán representables como fieles a un inicial hipotético constante. Pero, de manera interesante, la relación de consecuencia lógica no va en la dirección opuesta. Esto es, una agente puede manetener fidelidad a un inicial hipotético constante incluso si sus actualizaciones no siempre satisfacen Condicionalización. Por ejemplo, es posible que una agente tanto gane como pierda certezas entre dos momentos, t_i y t_j , y aun así exista un único inicial hipotético p_h tal que para cualquier $X \in \mathcal{L}$, $c_i(X) = p_h(X|E_i)$ y $c_j(X) = p_h(X|E_j)$.

Por ende, podemos obtener un modelo diacrónico plausible de agentes que tanto ganan como pierden certezas al generalizar Condicionalización, no para que exija que una agente condicione entre cada momento anterior t_i y posterior t_j , sino que en su lugar exija (sin importar qué ocurra con sus certezas) que ella ponga sus credibilidades en línea con un inicial hipotético constante a lo largo de su vida. Esta nueva norma diacrónica genera resultados plausibles para varias historias de olvido, tales como la presentada por Talbott. En casos en los que una agente estrictamente gana certezas entre dos momentos, imita los efectos de Condicionalización tradicional. Y en casos en los que una agente estrictamente pierde certezas entre momentos, nos da una Condicionalización temporal inversa. Esto es, las credibilidades incondicionales previas de la agente serán iguales a sus credibilidades posteriores condicionales sobre la información que perdió. Así, olvidar se vuelve algo similar a aprender a la inversa en el tiempo.

Desafortunadamente, pasar a esta nueva norma diacrónica no basta por sí mismo para enfrentar otro problema con Condicionalización: la manera en que trata información sensible al contexto. Aquí me referiré a afirmaciones “auto-localizantes” (*self-locating*) que cambian su valor de verdad a través de momentos, personas y ubicaciones –tales como “Hoy es martes”, “Yo soy un marinero” y “Nosotros estamos en Detroit”–. Para empezar,

48 Compárese con Schoenfield (2014) y Meacham (2016).

para modelar estas clases de afirmaciones en nuestro formalismo requeriremos añadir a nuestro lenguaje $\mathcal{L}$ algo como lo que Lewis (1979) llamó “proposiciones centradas”. Pero incluso entonces Condicionización enfrentará desafíos. Puede ser racional ahora mismo estar seguro de que es martes, pero esa certeza no seguirá siendo racional a perpetuidad.

El desafío de la sensibilidad al contexto se describe a veces como otro problema con la retención de certeza de Condicionización. Pero incluso cuando pasamos a una norma diacrónica que requiere fidelidad sólo con un inicial hipotético constante (y por ende permite pérdida de certeza) los problemas persisten. Esto se debe a que el sistema bayesiano fue diseñado para modelar agentes cuya evidencia cambió a través del tiempo, pero que usaron esa evidencia para evaluar hipótesis cuyos valores de verdad eran objetivos fijos.⁴⁹ Añadir en otro nivel de desplazamiento genera complicaciones para Condicionización, Condicionización de Jeffrey e iniciales hipotéticos.

Se han propuesto varios marcos conceptuales formales para modelar actualizaciones de credibilidad en proposiciones sensibles al contexto. Algunos conservan Condicionización, algunos usan iniciales hipotéticos, pero en cada nuevo caso se requieren normas adicionales para capturar el rango entero de fenómenos. No hay espacio para resumir las varias aproximaciones aquí.⁵⁰ Pero notaré que el problema de actualizar creencias auto-localizantes puede tener consecuencias importantes más allá de divertidos experimentos mentales filosóficos como el Problema de la Bella Durmiente (Elga, 2000). Por ejemplo, los argumentos de ajuste fino a favor de la existencia del multiverso y los debates sobre la interpretación apropiada de la mecánica cuántica, pueden ambos depender de cómo las agentes deberían gestionar credibilidades en proposiciones sensibles al contexto.⁵¹

49 En aplicaciones en filosofía de la ciencia, por ejemplo, las hipótesis científicas sobre las leyes físicas del universo o los orígenes evolutivos de los homínidos típicamente no cambian su valor de verdad a lo largo del tiempo.

50 Titelbaum (2016) proporciona un panorama a gran escala con numerosas referencias.

51 Para éstas y otras aplicaciones, véase Titelbaum (2013b).

Otros formalismos de la convicción

Para concluir, debería mencionar que hay varios formalismos alternativos para modelar los niveles de convicción variantes de la agente en afirmaciones. Primero, podemos pensar simplemente sobre si una agente tiene más convicción en una proposición que en otra. Integrar estas comparaciones produce un orden de convicción que puede flotar libre de cualesquiera asignaciones numéricas (véase Konek, 2019). Una segunda aproximación, llamada “teoría del *ranking*” (Spohn, 2012; Huber, 2019), vincula números al *ranking* de convicción, pero sólo funciona con la estructura de los enteros no negativos. Tercero, podemos emplear una estructura formal incluso más rica que los reales. Por ejemplo, en lugar de representar los niveles de convicción de una agente en un momento dado con una única distribución de probabilidad, podemos representarlos con un conjunto de tales distribuciones (Mahtani, 2019). O podemos tener una función con valores reales que rastree las actitudes de la agente y una separada (aunque relacionada) que rastree su evidencia. Esto produce una cuarta aproximación, comúnmente llamada “Teoría Dempster-Shafer” (Dempster, 1966; Shafer, 1976).

Cada una de estas aproximaciones puede ser apoyada por algunos de los tipos de argumentos arriba descritos, asimismo, cada una está plagada por algunos de los problemas de arriba también. Algunas permiten estructuras formales más flexibles y expresivas que el bayesianismo, mientras otras sacrifican poder expresivo en favor de plausibilidad psicológica. No intentaré elegir una favorita aquí. Pero vale la pena notar que, de entre todos

los formalismos para representar niveles de convicción diversos, ninguno se estudia o se aplica más en la actualidad que la aproximación credal con valores reales.⁵²

52 Gracias a Richard Pettigrew y Jonathan Weisberg, especialmente a este último por comentarios detallados y múltiples sugerencias de citas. Mucho del material en esta sección ha sido adaptado de mi libro (Titelbaum, 2022), el cual cubre casi todos los temas presentados aquí con mucho mayor profundidad.

Apéndice

He aquí el esbozo de una demostración para el teorema de representación mencionado en la página 53. Asumiremos que, en la teoría de la decisión de interés, lo siguiente se sostiene.

- Axiomas estructurales que aseguran que los actos de apuestas con varias estructuras (como se describen en la demostración abajo) siempre están disponibles para la agente.
- Principio de dominancia débil: cuando los actos son independientes de estados, si no hay un estado en que el acto A produce una mayor utilidad que el acto B , entonces A no se prefiere sobre B .
- Principio de dominancia fuerte: cuando los actos son independientes de estados, si el acto A produce una mayor utilidad que el acto B en todo estado, entonces A se prefiere sobre B .
- Para cualesquiera actos A y B , la agente prefiere A sobre B sólo en caso de que $UE(A) > UE(B)$, donde UE se calcula tal como se describe en el texto principal.

Los principios de dominancia descritos arriba emplean una noción de independencia acto/estado, la noción relevante variará dependiendo de cuál teoría de la decisión (evidencial, causal, etc.) esté en juego. Así que desplegar los detalles de la demostración de abajo para una teoría de la decisión específica requerirá mostrar que los actos y estados que aparecen en cada paso de la demostración son independientes en el sentido relevante. Dados los tipos de actos involucrados, eso debería ser bastante directo.

Nota que lo siguiente es un corolario del principio de dominancia débil.

- Principio de equivalencia: cuando los actos son independientes de estados, si dos actos producen la misma utilidad uno que otro en todo estado posible, la agente es indiferente entre ellos.

El argumento es simplemente que, si A y B producen la misma utilidad en todo estado posible, entonces por dominancia débil A no se prefiere sobre B y B no se prefiere sobre A . Así que la agente es indiferente entre ellos y $UE(A) = UE(B)$.

Para mostrar que cualquier función de credibilidad c que aparezca en una teoría de la decisión con los rasgos arriba enlistados debe ser una transformación escalar positiva de una función de probabilidad, necesitamos demostrar que satisface cuatro condiciones.

1. Cualquier tautología en $\mathcal{L}$ recibe el mismo valor- c .

Demostración. Supóngase para *reductio* que tenemos dos tautologías $\top_1, \top_2 \in \mathcal{L}$ tales que la agente asigna una credibilidad de x a la primera y una credibilidad de y distinta a la segunda. Considera un acto que paga 1 útil en $\top_1$ y 0 útiles en caso contrario, y un acto que paga 1 útil en $\top_2$ y 0 útiles en caso contrario. La agente asignará al primer acto una utilidad esperada de x , al segundo acto una utilidad esperada de y . Puesto que x y y son diferentes, la agente preferirá un acto sobre otro. Pero los dos actos producen cada uno la misma ganancia (1 útil) en cualquier estado posible, así que hemos violado el principio de equivalencia (y, por ende, de dominancia débil).

2. Para cualesquiera tautología y contradicción $\top, \perp \in \mathcal{L}$, $c(\top) > c(\perp)$.

Demostración. Supóngase para *reductio* que tenemos $\top$ y $\perp$ tales que $c(\top) \leq c(\perp)$. Ahora considera un acto que paga 1 útil en $\top$ y 0 útiles en caso contrario, y otro acto que paga 1 útil en $\perp$ y 0 en caso contrario. Dada la suposición, la agente asignará al primer acto una utilidad esperada no mayor a la del segundo. Pero el primer acto produce una mayor utilidad que el segundo en cual-

quier estado posible, así que por dominancia fuerte el primer acto debe recibir una utilidad esperada más alta.

3. Para cualesquiera $X, Y \in \mathcal{L}$ mutuamente excluyentes $c(X \vee Y) = c(X) + c(Y)$.

Demostración. Primero considera el acto de comprar una apuesta que paga 1 útil en X , 1 útil en Y y 0 útiles en cualquier otro caso. Puesto que X y Y son mutuamente excluyentes, podemos particionar los estados posibles en X , Y , y $\neg X \& \neg Y$. Usando esta partición, la utilidad esperada de este acto es

$$\begin{aligned} c(X) \cdot u(X) + c(Y) \cdot u(Y) + c(\neg X \& \neg Y) \cdot u(\neg X \& \neg Y) = \\ c(X) \cdot 1 + c(Y) \cdot 1 + c(\neg X \& \neg Y) \cdot 0 = c(X) + c(Y). \end{aligned} \quad (27)$$

Ahora considera el acto de comprar una apuesta que paga 1 útil en $X \vee Y$ y 0 útiles en caso contrario. Al particionar los estados en $X \vee Y$ y $\neg(X \vee Y)$, la utilidad esperada de este acto es

$$\begin{aligned} c(X \vee Y) \cdot u(X \vee Y) + c(\neg[X \vee Y]) \cdot u(\neg[X \vee Y]) = \\ c(X \vee Y) \cdot 1 + c(\neg[X \vee Y]) \cdot 0 = c(X \vee Y). \end{aligned} \quad (28)$$

Estos dos actos dan la misma ganancia en cualquier estado posible, así que para satisfacer el principio de equivalencia la agente debe ser indiferente entre ellos. Esto significa que sus utilidades esperadas son iguales, así que $c(X \vee Y) = c(X) + c(Y)$.

4. Para cualquier $X \in \mathcal{L}$, $c(X) \geq 0$.

Demostración. Primero, mostramos que no puede haber $T, \top \in \mathcal{L}$ tales que $\top$ es una tautología y $c(T) > c(\top)$. Supóngase para *reductio* que tuviésemos tales dos proposiciones T y $\top$. Ahora considera un acto que paga 1 útil si T es verdadera y 0 útiles en caso contrario, y un acto que paga 1 útil si $\top$ es verdadera y 0 útiles en caso contrario. El primer acto tiene una utilidad esperada $c(T)$, mientras el segundo tiene una utilidad esperada $c(\top)$. Por nuestra

suposición, la agente prefiere el primer acto. Pero puesto que $\top$ es verdadera en todo estado, no hay ningún estado en el que el primer acto produzca una mayor utilidad que el segundo. Así que hemos violado dominancia débil.

Ahora al resultado principal. Asúmase para *reductio* que existe un $X \in \mathcal{L}$, tal que $c(X) < 0$. Puesto que X y $\neg X$ son mutuamente excluyentes, $c(X \vee \neg X) = c(X) + c(\neg X)$ por (3) arriba. Si $c(X) < 0$, entonces $c(X \vee \neg X) < c(\neg X)$. Pero $X \vee \neg X$ es una tautología, así que esto es imposible.

Referencias

- Adams, E. (1962). On rational betting systems. *Archiv für mathematische Logik und Grundlagenforschung*, 6, 7–29.
- Allais, M. (1953). Le Comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine. *Econometrica*, 21, 503–46.
- Armendt, B. (1980). Is there a Dutch Book argument for probability kinematics? *Philosophy of Science*, 47, 583–588.
- Armendt, B. (1992). Dutch strategies for diachronic rules: When believers see the sure loss coming. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1, 217–229.
- Arntzenius, F. (2003). Some problems for conditionalization and reflection. *The Journal of Philosophy*, 100, 356–370.
- Berker, S. (2013). Epistemic teleology and the separateness of propositions. *Philosophical Review*, 122, 337–93.
- Bernoulli, J. (1713). *Ars conjectandi*. Basiliae.
- Bjerring, J. C. y Skipper, M. (2022). Bayesianism for non-ideal agents. *Erkenntnis*, 87, 93–115.
- Boole, G. (1952). On the application of the theory of probabilities to the question of the combination of testimonies or judgments. *Studies in Logic and Probability*, 308–85.
- Bovens, L. y Hartmann, S. (2003). *Bayesian epistemology*. Oxford: Oxford University Press.
- Bradley, R. (2007). Reaching a consensus. *Social Choice and Welfare*, 29, 609–32.
- Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78, 1–3.
- Briggs, R. A. (2009). Distorted reflection. *The Philosophical Review*, 118, 59–85.

- Briggs, R. A. (2019). Conditionals. En R. Pettigrew y J. Weisberg (eds.) (2019). *Open handbook of formal epistemology* (pp. 543-590). Ontario: PhilPapers Foundation.
- Briggs, R. A. y Pettigrew, R. (2020). An accuracy-dominance argument for Conditionalization. *Noûs*, 54(1), 162-181.
- Buchak, L. (2013). *Risk and rationality*. Oxford: Oxford University Press.
- Carnap, R. (1950). *Logical foundations of probability*. Chicago: University of Chicago Press.
- Carnap, R. (1962). *Logical foundations of probability* (2nd). Chicago: University of Chicago Press.
- Carr, J. R. (2015). Don't stop believing. *Canadian Journal of Philosophy*, 45.
- Carr, J. R. (2017). Epistemic utility theory and the aim of belief. *Philosophy and Phenomenological Research*, 95, 511-34.
- Chihara, C. (1981). Quine and the confirmational paradoxes. En P. French, H. Wettstein, y T. Uehling (Eds.), *Midwest studies in philosophy 6: Foundations of analytic philosophy* (pp. 425-52). University of Minnesota Press.
- Christensen, D. (1999). Measuring confirmation. *The Journal of Philosophy*, 96, 437-61.
- Christensen, D. (2004). *Putting logic in its place*. Oxford: Oxford University Press.
- Christensen, D. (2009). Disagreement as evidence: The epistemology of controversy. *Philosophy Compass*, 4, 756-67.
- Christensen, D. y Lackey, J. (Eds.). (2013). *The epistemology of disagreement: New essays*. Oxford: Oxford University Press.
- Crupi, V. (2020). Confirmation. En E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2020). Metaphysics Research Lab, Stanford University.
- Davies, M. (1998). Externalsism, architecturalism, and epistemic warrant. En C. Wright, B. Smith y C. Macdonald (Eds.), *Knowing our own minds* (pp. 321-61). Oxford: Oxford University Press.
- De Finetti, B. (1937/1964). Foresight: Its logical laws, its subjective sources. En H. E. Kyburg Jr y H. Smokler (Eds.), *Studies in subjective probability* (pp. 94-158). Originalmente publicado como "La prévision; ses lois logiques, ses sources subjectives" en *Annales de l'Institut Henri Poincaré, Volume 7*, 1-68.
- De Finetti, B. (1974). *Theory of probability*. New York: Wiley.

- Dempster, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *Annals of Mathematical Statistics*, 37, 355–74.
- Douven, I. (2012). The Lottery Paradox and the pragmatics of belief. *Dialectica*, 66, 351–73.
- Dretske, F. (1970/2023). Operadores epistémicos (Alejandro Nogara, trad.) *Disputatio*, 26 (12): 71-89.
- Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: The MIT Press.
- Easwaran, K. (2013). Expected accuracy supports conditionalization—and conglomerability and reflection. *Philosophy of Science*, 80, 119–142.
- Easwaran, K. (2014). Regularity and hyperreal credences. *Philosophical Review*, 123, 1–41.
- Easwaran, K. (2019). Conditional probabilities. En R. Pettigrew y J. Weisberg (eds.) (2019). *Open handbook of formal epistemology* (pp. 131-198). Ontario: PhilPapers Foundation.
- Easwaran, K., Fenton-Glynn, L., Hitchcock, C. y Velasco, J. D. (2016). Updating on the credences of others: Disagreement, agreement, and synergy. *Philosophers' Imprint*, 16, 1–39.
- Eells, E. (1982). *Rational decision and causality*. Cambridge Studies in Philosophy. Cambridge: Cambridge University Press.
- Eells, E. (1985). Problems of old evidence. *Pacific Philosophical Quarterly*, 66, 283–302.
- Elga, A. (2000). Self-locating belief and the Sleeping Beauty problem. *Analysis*, 60, 143–7.
- Elga, A. (2007). Reflection and disagreement. *Noûs*, 41, 478–502.
- Feldman, R. (2007). Reasonable religious disagreements. En L. M. Antony (Ed.), *Philosophers without gods: Meditations on atheism and the secular life*. Oxford: Oxford University Press.
- Fitelson, B. (2008). A decision procedure for probability calculus with applications. *The Review of Symbolic Logic*, 1, 111–125.
- Fitelson, B. (2015). The strongest possible Lewisian triviality result. *Thought*, 4, 69–74.
- Fitelson, B. y Hawthorne, J. (2010). How Bayesian confirmation theory handles the Paradox of the Ravens. *Boston Studies in the Philosophy of Science*, 284.
- Foley, R. (1993). *Working without a net*. Oxford: Oxford University Press.

- Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory. En J. Earman (Ed.), *Testing scientific theories* (Vol. 10, pp. 99–132). *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press.
- Gibbard, A. y Harper, W. (1978). Counterfactuals and two kinds of expected utility. En C. A. Hooker, J. L. Leach y E. F. McClennan (Eds.), *Foundations and applications of decision theory* (Vol. 13a, pp. 125–62). University of Western Ontario Series in Philosophy of Science. Dordrecht: D. Reidel Publishing Company.
- Glymour, C. (1980). *Theory and evidence*. Princeton, NJ: Princeton University Press.
- Goodman, N. (1955/2004). *Hecho, ficción y pronóstico*. (Jorge Rodríguez Marqueze, trad.). Madrid: Síntesis.
- Greaves, H. (2013). Epistemic decision theory. *Mind*, 122, 915–52.
- Greaves, H. y Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115, 607–632.
- Hacking, I. (1967). Slightly more realistic personal probability. *Philosophy of Science*, 34, 311–325.
- Hájek, A. (2003). What conditional probability could not be. *Synthese*, 137, 273–323.
- Hájek, A. (2007). The reference class problem is your problem too. *Synthese*, 156, 563–85.
- Hájek, A. (2009). Arguments for –or against– probabilism? En F. Huber y C. Schmidt-Petri (Eds.), *Degrees of belief* (Vol. 342, pp. 229–251). Synthese Library. Springer.
- Hájek, A. (2011). Triviality pursuit. *Topoi*, 30, 3–15.
- Harman, G. (1986). *Change in view*. Boston: The MIT Press.
- Hawthorne, J. y Fitelson, B. (2004). Re-solving irrelevant conjunction with probabilistic independence. *Philosophy of Science*, 71, 505–514.
- Hempel, C. G. (1945/1979). Estudios sobre la lógica de la confirmación. En *La explicación científica. Estudios sobre la filosofía de la ciencia*. (pp. 13-59) (M. Frassinetti de Gallo, Néstor Míguez e Irma Ruiz Aused, trads.). Barcelona: Paidós.
- Henderson, L. (2013). Bayesianism and inference to the best explanation. *British Journal for the Philosophy of Science*, 65, 687–715.
- Hintikka, J. (1975). Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4, 475–84.

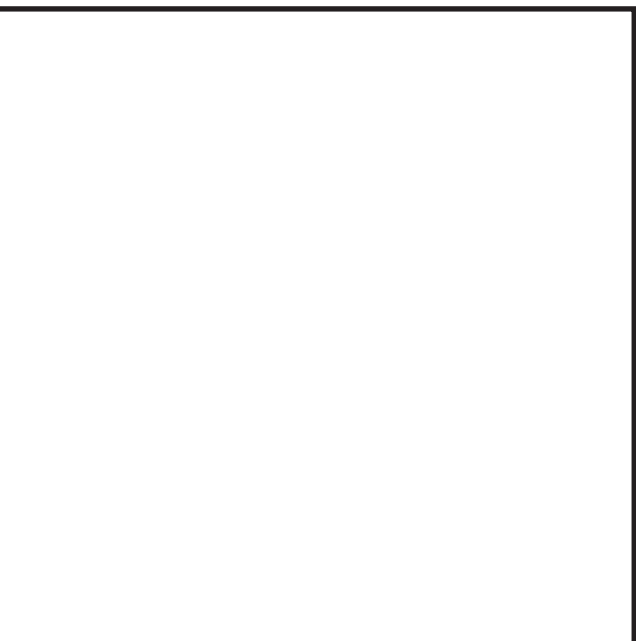
- Hitchcock, C. R. (2012). Probabilistic causation. En E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2012).
- Holton, R. (2014). Intention as a model for belief. En M. Vargas y G. Yaffe (Eds.), *Rational and social agency: The philosophy of Michael Bratman* (pp. 12–37). Oxford: Oxford University Press.
- Hosiasson-Lindenbaum, J. (1940). On confirmation. *The Journal of Symbolic Logic*, 5, 133–48.
- Howson, C. (1992). Dutch Book arguments and consistency. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2, 161–8.
- Howson, C. y Urbach, P. (2006). *Scientific reasoning: The Bayesian approach* (3rd). Chicago: Open Court.
- Huber, F. (2019). Ranking theory. En R. Pettigrew y J. Weisberg (eds.) (2019). *Open handbook of formal epistemology* (pp. 397–436). Ontario: PhilPapers Foundation.
- Huemer, M. (2011). Does probability theory refute coherentism? *Journal of Philosophy*, 108, 463–72.
- Jaynes, E. T. (1957a). Information theory and statistical mechanics I. *Physical Review*, 106, 620–30.
- Jaynes, E. T. (1957b). Information theory and statistical mechanics II. *Physical Review*, 108, 171–90.
- Jeffrey, R. C. (1965). *The logic of decision* (1st). McGraw-Hill series in probability and statistics. New York: McGraw-Hill.
- Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65, 575–603.
- Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge: Cambridge University Press.
- Joyce, J. M. (2009). Accuracy and coherence: Prospects for an alethic epistemology of partial belief. En F. Huber y C. Schmidt-Petri (Eds.), *Degrees of belief* (Vol. 342, pp. 263–297). Synthese Library. Springer.
- Kemeny, J. G. (1955). Fair bets and inductive probabilities. *The Journal of Symbolic Logic*, 20, 263–273.
- Kolmogorov, A. N. (1933/1950). *Foundations of the theory of probability*. Translation edited by Nathan Morrison. New York: Chelsea Publishing Company.
- Konek, J. (2019). Comparative probabilities. En R. Pettigrew y J. Weisberg (eds.) (2019). *Open handbook of formal epistemology* (pp. 267–348). Ontario: PhilPapers Foundation.
- Kyburg, H. E., Jr. (1961). *Probability and the logic of rational belief*. Middletown: Wesleyan University Press.

- Laplace, P. S. (1814/1985). *Ensayo filosófico de la teoría de probabilidades*. (Pilar Castrillo, trad.). Madrid: Alianza Editorial.
- Leitgeb, H. (2017). *The stability of belief: How rational belief coheres with probability*. Oxford: Oxford University Press.
- Levi, I. (1987). The demons of decision. *The Monist*, 70, 193–211.
- Lewis, D. (1971). Immodest inductive methods. *Philosophy of Science*, 38, 54–63.
- Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85, 297–315.
- Lewis, D. (1979). Attitudes *de dicto* and *de se*. *The Philosophical Review*, 88, 513–543.
- Lewis, D. (1980). A subjectivist's guide to objective chance. En R. C. Jeffrey (Ed.), *Studies in inductive logic and probability* (Vol. 2, pp. 263–294). Berkeley: University of California Press.
- Lewis, D. (1981). Causal decision theory. *Australasian Journal of Philosophy*, 59, 5–30.
- Lin, H. y Kelly, K. T. (2012). A geo-logical solution to the Lottery Paradox. *Synthese*, 186, 531–75.
- Lipton, P. (2004). *Inference to the best explanation* (2nd). London: Routledge.
- Littlejohn, C. (2018). Stop making sense? On a puzzle about rationality. *Philosophy and Phenomenological Research*, 96(2), 257–272.
- Locke, J. (1689/1999). *Ensayo sobre el entendimiento humano*. (Edmundo O'Gorman, trad.). México: Fondo de Cultura Económica.
- Maher, P. (1996). Subjective and objective confirmation. *Philosophy of Science*, 63, 149–74.
- Maher, P. (2002). Joyce's argument for probabilism. *Philosophy of Science*, 96, 73–81.
- Mahtani, A. (2019). Imprecise probabilities. En R. Pettigrew y J. Weisberg (eds.) (2019). *Open handbook of formal epistemology* (pp. 107–130). Ontario: PhilPapers Foundation.
- Makinson, D. C. (1965). The paradox of the preface. *Analysis*, 25, 205–7.
- Meacham, C. J. (2010). Two mistakes regarding the Principal Principle. *British Journal for the Philosophy of Science*, 61, 407–31.
- Meacham, C. J. (2016). Ur-priors, conditionalization, and ur-prior conditionalization. *Ergo*, 3(17), 444–492.
- Meacham, C. J. y Weisberg, J. (2011). Representation theorems and the foundations of decision theory. *Australasian Journal of Philosophy*, 89, 641–663.

- Moss, S. (2018). *Probabilistic knowledge*. Oxford: Oxford University Press.
- Okasha, S. (2000). Van Fraassen's critique of inference to the best explanation. *Studies in History and Philosophy of Science*, 691–710.
- Olsson, E. (2017). Coherentist theories of epistemic justification. En E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2017). Metaphysics Research Lab, Stanford University.
- Pettigrew, R. (2013). A new epistemic utility argument for the Principal Principle. *Episteme*, 10, 19–35.
- Pettigrew, R. (2014). Accuracy, risk, and the Principle of Indifference. *Philosophy and Phenomenological Research*, 90.
- Pettigrew, R. (2016). *Accuracy and the laws of credence*. Oxford: Oxford University Press.
- Predd, J., Seiringer, R., Lieb, E., Osherson, D., Poor, V., y Kulkarni, S. (2009). Probabilistic coherence and proper scoring rules. *IEEE Transactions on Information Theory*, 55, 4786–4792.
- Ramsey, F. P. (1926/2005). Verdad y probabilidad. En *Obra filosófica completa*. (pp. 156–198) (María José Frápolli, Francesc Camós y Javier Rodríguez Alcázar, trads.). Granada: Editorial Comares.
- Rasmussen, M. y Steglich-Petersen, A. (eds.). (2019). *Higher-order evidence: New essays*. Oxford: Oxford University Press.
- Reichenbach, H. (1956). The principle of common cause. En *The direction of time* (pp. 157–160). University of California Press.
- Rosenkrantz, R. (1977). *The paradoxes of confirmation*. Synthese Library. Dordrecht: D. Reidel Publishing Company.
- Rosenkrantz, R. (1981). *Foundations and applications of inductive probability*. Atascadero, CA: Ridgeview Press.
- Russell, J., Hawthorne, J. y Buchak, L. (2015). Groupthink. *Philosophical Studies*, 172, 1287–1309.
- Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
- Savage, L. J. (1967). Difficulties in the theory of personal probability. *Philosophy of Science*, 34, 305–310.
- Schoenfeld, M. (2014). Permission to believe: Why permissivism is true and what it tells us about irrelevant influences on belief. *Noûs*, 48, 193–218.
- Seidenfeld, T. (1986). Entropy and uncertainty. *Philosophy of Science*, 53, 467–491.
- Seidenfeld, T., Schervish, M. J. y Kadane, J. (2017). Non-conglomerability for countably additive measures that are not k-additive. *The Review of Symbolic Logic*, 10(2), 284–300.

- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.
- Shimony, A. (1955). Coherence and the axioms of confirmation. *Journal of Symbolic Logic*, 20, 1–28.
- Shimony, A. (1988). An Adamite derivation of the calculus of probability. En J. Fetzer (Ed.), *Probability and causality* (pp. 151–161). Dordrecht: Reidel.
- Shogenji, T. (1999). Is coherence truth-conducive? *Analysis*, 59, 338–45.
- Skyrms, B. (1987a). Coherence. En N. Rescher (Ed.), *Scientific inquiry in philosophical perspective* (pp. 225–42). Pittsburgh: University of Pittsburgh Press.
- Skyrms, B. (1987b). Dynamic coherence and probability kinematics. *Philosophy of Science*, 54, 1–20.
- Smithies, D. (2015). Ideal rationality and logical omniscience. *Synthese*, 192, 2769–93.
- Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications*. Oxford: Oxford University Press.
- Suppes, P. (1974). *Probabilistic metaphysics*. Uppsala: University of Uppsala Press.
- Swinburne, R. (1971). The paradoxes of confirmation: A survey. *American Philosophical Quarterly*, 8, 318–30.
- Talbott, W.J. (1991). Two principles of Bayesian epistemology. *Philosophical Studies*, 62, 135–150.
- Teller, P. (1973). Conditionalization and observation. *Synthese*, 26, 218–258.
- Thoma, J. (2019). Decision theory. En R. Pettigrew y J. Weisberg (eds.) (2019). *Open handbook of formal epistemology* (pp. 57–106). Ontario: PhilPapers Foundation.
- Titelbaum, M. G. (2013a). *Quitting certainties: A Bayesian framework modeling degrees of belief*. Oxford: Oxford University Press.
- Titelbaum, M. G. (2013b). Ten reasons to care about the Sleeping Beauty Problem. *Philosophy Compass*, 8, 1003–17.
- Titelbaum, M. G. (2015). Rationality's fixed point (or: In defense of right reason). En T. S. Gendler y J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 5, pp. 253–94). Oxford University Press.
- Titelbaum, M. G. (2016). Self-locating credences. En A. Hájek y C. R. Hitchcock (Eds.), *The Oxford handbook of probability and philosophy* (pp. 666–680). Oxford: Oxford University Press.

- Titelbaum, M. G. (2022). *Fundamentals of Bayesian epistemology*. 2 vols. Oxford: Oxford University Press.
- van Fraassen, B. C. (1983). Calibration: A frequency justification for personal probability. En R. Cohen y L. Laudan (Eds.), *Physics, philosophy, and psychoanalysis* (pp. 295–319). Dordrecht: Reidel.
- van Fraassen, B. C. (1984). Belief and the will. *The Journal of Philosophy*, 81, 235–256.
- van Fraassen, B. C. (1989). *Laws and symmetry*. Oxford: Clarendon Press.
- von Mises, R. (1928/1946). *Probabilidad, estadística y verdad*. (Juan Carlos Grimberg, trad. Edición castellana del original en alemán *Wahrscheinlichkeit, Statistik und Wahrheit*). Buenos Aires: Espasa-Calpe.
- von Neumann, J. y Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd). Princeton, NJ: Princeton University Press.
- Weirich, P. (2012). Causal decision theory. En E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2012).
- Weisberg, J. (2007). Conditionalization, reflection, and self-knowledge. *Philosophical Studies*, 135, 179–197.
- Weisberg, J. (2009). Locating IBE in the Bayesian framework. *Synthese*, 167, 125–44.
- Wenmackers, S. (2019). Infinitesimal probabilities. En R. Pettigrew y J. Weisberg (eds.) (2019). *Open handbook of formal epistemology* (pp. 199–265). Ontario: PhilPapers Foundation.
- White, R. (2005). Epistemic permissiveness. *Philosophical Perspectives*, 19, 445–459.
- Williamson, T. (2007). How probable is an infinite sequence of heads? *Analysis*, 67, 173–80.
- Wright, C. (2003). Some reflections on the acquisition of warrant by inference. En S. Nuccetelli (Ed.), *New essays on semantic externalism*. Cambridge, MA: MIT Press.
- Zynda, L. (1995). Old evidence and new theories. *Philosophical Studies*, 77, 67–95.
- Zynda, L. (2000). Representation theorems and realism about degrees of belief. *Philosophy of Science*, 67, 45–69.



Epistemología bayesiana

Primera edición 2024

(versión electrónica)

El cuidado y diseño de la edición estuvieron
a cargo del Departamento Editorial
de la Dirección General de Difusión y Vinculación
de la Universidad Autónoma de Aguascalientes.

